



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Exploration of a method to validate surrogate safety measures with a focus on vulnerable road users

Laureshyn, Aliaksei; Johnsson, Carl; Madsen, Tanja Kidholm Osmann; Várhelyi, András; de Goede, Maartje; Svensson, Åse; Saunier, Nicolas; van Haperen, Wouter

Published in:

Proceedings of the Road Safety & Simulation International Conference, 17-19 October 2017

Publication date:
2017

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Laureshyn, A., Johnsson, C., Madsen, T. K. O., Várhelyi, A., de Goede, M., Svensson, Å., Saunier, N., & van Haperen, W. (2017). Exploration of a method to validate surrogate safety measures with a focus on vulnerable road users. In *Proceedings of the Road Safety & Simulation International Conference, 17-19 October 2017*

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 ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



LUND UNIVERSITY

Exploration of a method to validate surrogate safety measures with a focus on vulnerable road users

Laureshyn, Aliaksei; Johnsson, Carl; Madsen, Tanja; Varhelyi, Andras; de Goede, Maartje; Svensson, Åse; Saunier, Nicolas; van Haperen, Wouter

Published in:

Proceedings of the Road Safety & Simulation International Conference, 17-19 October 2017

2017

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Laureshyn, A., Johnsson, C., Madsen, T., Varhelyi, A., de Goede, M., Svensson, Å., Saunier, N., & van Haperen, W. (2017). Exploration of a method to validate surrogate safety measures with a focus on vulnerable road users. In *Proceedings of the Road Safety & Simulation International Conference, 17-19 October 2017* [43]

Total number of authors:

8

General rights

Unless other specific re-use rights are stated the following general rights apply:

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

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

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.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Exploration of a method to validate surrogate safety measures with a focus on vulnerable road users

Aliaksei Lareshyn¹, Carl Jonsson², Tanja Kidholm Osmann Madsen³, András Várhelyi⁴,
Maartje de Goede⁵, Åse Svensson⁶, Nicolas Saunier⁷, Wouter van Haperen⁸

¹ Dept. of Technology & Society, Faculty of Engineering, LTH, Lund University
Box 118, 22100 Lund, Sweden, aliaksei.lareshyn@tft.lth.se

² Dept. of Technology & Society, Faculty of Engineering, LTH, Lund University
Box 118, 22100 Lund, Sweden, carl.jonsson@tft.lth.se

³ Aalborg University, Department of Civil Engineering, Traffic Research Group
Thomas Manns Vej 23, DK-9220 Aalborg, Denmark, tkom@civil.aau.dk

⁴ Dept. of Technology & Society, Faculty of Engineering, LTH, Lund University
Box 118, 22100 Lund, Sweden, andras.varhelyi@tft.lth.se

⁵ TNO, Dept. of Perceptual and Cognitive Systems
3769 DE, Soesterberg, The Netherlands, maartje.degoede@tno.nl

⁶ Dept. of Technology & Society, Faculty of Engineering, LTH, Lund University
Box 118, 22100 Lund, Sweden, ase.svensson@tft.lth.se

⁷ Dept. of Civil, Geological and Mining Engineering, Polytechnique Montréal
16 H3C 3A7 Montréal, Québec, Canada, nicolas.saunier@polymtl.ca

⁸ Transportation Research Institute (IMOB), Hasselt University
Wetenschapspark 5 bus 6, B-3590 Diepenbeek, Belgium, wouter.vanhaperen@uhasselt.be

Abstract

Background. Traditional crash-based analysis of road safety at individual sites has its shortcomings due to low numbers and the random nature of crashes at individual sites and the related statistical issues, as well as the under-reporting of crashes and lack of information on contributing factors and the process preceding crashes. To get around the problem, road safety analysis based on surrogate measures of safety, i.e. not based on crashes, can be used. However, the question whether surrogate measures are valid indicators for safety remains unanswered and only a few attempts have actually been made to carry out proper large-scale validation studies.

Aim. This work presents a methodological approach for a large-scale validation study of surrogate safety indicators focusing on vulnerable road users. With only one site analyzed so far, it presents the exploration of the data and of the performance of the technical tools used in the study.

Method. Video-filming and consequent video analysis are used to measure the surrogate safety indicators. In the first step, the video is “condensed” using a watchdog software RUBA that selects situations with an encounter of a cyclist or pedestrian and a motor vehicle. At a later stage, the trajectories of the individual road users are produced using a semi-automated tool T-Analyst and several surrogate safety indicators are tested to set a severity score for an encounter. The performance of the surrogate indicators will be compared to the expected number of accidents at each site and availability of the data for developing a safety performance function (SPF) that is country-, manoeuvre- and type of VRU-specific are explored.

Results & Conclusion. From methodological perspective, limited accident data available seriously complicates building a reliable SPF (“ground truth”) against which the surrogate safety measures could be validated; some other, “indirect” methods of validation might be required. We present also the performance of the software tools and applicability of the various surrogate safety indicators that were tested.

Keywords

Road Safety; Surrogate safety measures; Validation; Vulnerable Road Users

1 Background

The limitations of the traditional crash-based analysis have been repeatedly described in the literature and include the random nature and low crash counts at individual sites and the related statistical issues, the under-reporting of crashes and lack of information describing the process and contributing factors to crashes [1-4]. The surrogate methods, on the opposite, enable quick, pro-active and detail-rich safety evaluations by studying traffic events of lower severity and thus much more frequent compared to crashes [5].

The surrogate safety methods rest on the theoretical assumption of continuity among elementary events in traffic, often illustrated with the “safety pyramid” (see Figure 1). The very top of the pyramid represents fatal and injury crashes that are the most severe and most rare events in the whole continuum. The bottom of the pyramid, on the other hand, represents the “normal” traffic, which takes place most of the time. If the nature of the relation between the pyramid “layers” is known, it is possible to estimate the frequency of the rare events (fatal/serious injury crashes) based on observation of less severe and more frequent events.

The term “severity” has been often interpreted as “nearness to a crash as such”. We strongly argue that the more correct definition would be “nearness to a personal injury” [6], which is based on the Vision Zero philosophy prioritizing elimination of fatal and serious injury crashes over any other crashes with less severe outcomes [7]. Both definitions, though quite intuitive, require operationalization in more objective terms to make possible the actual construction of the safety pyramid. It is obvious, that the choice of the indicator(s) used to quantify the severity has a direct effect on the final pyramid shape and also on the position of an individual event in the safety hierarchy (which is the reason for why it is often so difficult to compare results from different surrogate safety

studies as they are simply “not compatible” - [8]). Very often, the severity of an event is described by “nearness” of the two road users, either in space [9] or time [10, 11] which can roughly be interpreted as a measure of probability for a collision. Another important aspect, however, is the potential outcome of the hypothetical collision, i.e. how likely it is to result in serious injuries of the involved road users. This becomes particularly important in case the situation involves vulnerable road users (VRUs) that are much likely to be injured compared to motor vehicle occupants, even in case of a minor collision. Ideally, these two dimensions should be “weighed” together into a single severity measure. This has been implemented in some of the traffic conflict techniques by setting arbitrary rules for which severity score should be assigned to events based on collision risk and its potential outcomes [12, 13]. Recently, some objective measures have also been suggested [6, 14-16].

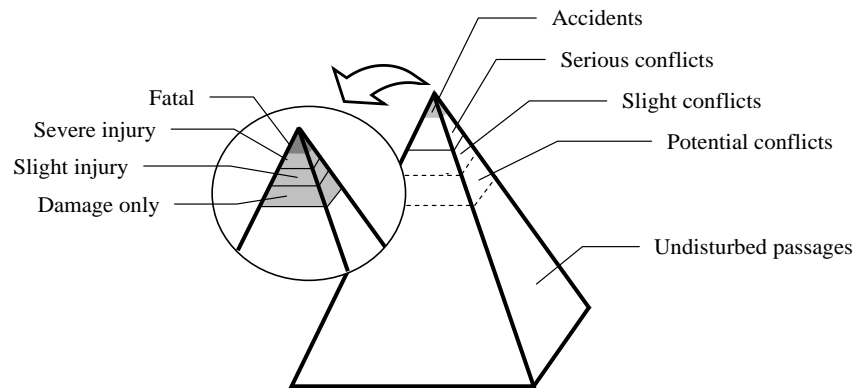


Figure 1. “The safety pyramid” (adopted from [17]).

The two very important characteristics of any surrogate safety methodology are its reliability and validity [1]. **Reliability** refers to the accuracy of the measurement tools that should remain within the same limits regardless of the study site, light, weather and traffic conditions, etc. Thus, the reliability ensures that the differences in the measured values are due to actual difference of the underlying parameter of interest and not due to the tool’s inadequacy. The use of human observers as the main “tool” in many of the original traffic conflict techniques has been constantly questioned and tested as their judgements were expected to be influenced by the emotional state, fatigue, loss of attention, etc. [17-19]. The development of numerous automated tools during the last decade [1, 20-27] successfully addresses many of the natural weaknesses of human observers. Though very complex technically for the moment, some tracking tools reach nearly 100% accuracy in detection and tracking of all road users [28] which can be taken as a very promising indication that the reliability challenge for the surrogate safety methodologies will not be a major issue in a very near future.

The **validity** refers to the ability of the chosen indicator(s) to actually reflect the desired quality we want to measure. In case of the surrogate safety analysis it means a robust relation of the surrogate measure with the personal injury crashes expected to occur at the studied site. It is important to understand that the “absolute validity” for any method is an idealistic concept that must be pursued, but might be very hard to reach (if ever) [29]. Thus, different degrees and types of validity might be acceptable in surrogate safety analysis:

- **Absolute product validity** – the ideal case when the expected number of crashes can be inferred from the surrogate safety measure, thus allowing to report “traditional” safety indicators like injury crash rates, number of lives saved, to make cost-benefit analyses, etc.
- **Relative product validity** – if the surrogate measure can indicate the direction of change (improvement or deterioration) in safety, but not its absolute extent (expressed as the change in expected number of injury crashes).
- **Process validity** – similarity in how the real crashes and the surrogate safety events evolve. While it has nothing to do with measuring the safety level, much can be learnt about the crash contributing factors that otherwise are seldom available in the crash reports and even in-depth studies.

There has been very few validation studies that actually compared the observed surrogate safety measures and some kind of accident data (for a comprehensive review of these efforts we address the reader to [30]). The approaches to product validation suggested in the literature can be classified as:

- Analysis of the linear correlation between the numbers of observed safety critical events and recorded crashes [31]. It turned out quite early that since both crash counts and safety critical events are subjects to random variation, the correlation (or its absence) might be often misleading if only few data points are available. Moreover, as the traffic safety improved, the crash counts became very low and similar to categorical data (“0 crashes” or “1 crash” or “2 crashes”) for which the liner regression methods are not really suitable.

- Analysis of the linear correlation between the number of observed safety critical events and the expected number of crashes provided by a safety performance function (SPF) (having traffic flow as input) [32].
- Analysis of the variance of the ratio between the estimated expected number of crashes and observed number of (or estimated expected number of) safety critical events [33, 34].
- Analysis of safety performance functions that estimate the expected number of crashes from the number of safety critical events (and not flow) [35]. In this case, the relation is not expected to be linear and different functional shapes can be tested.
- Comparison of the expected number of crashes (based on crash history) with the expected number of crashes calculated using the extreme value theory methods [36].
- Comparison of conclusions from before-after studies based on frequency/location of the safety critical events and based on crash history when it becomes available [37].
- Analysis of the similarities of the safety performance functions estimating the crash number and critical event number based on the traffic flow [38].

2 Aim

This work presents a methodological approach and first results of a large scale validation study of surrogate safety indicators focusing on vulnerable road users [39]. Primarily, we focus on the “product validity” (i.e. ability of a surrogate measure to estimate the “expected number of accidents”). While the full study will cover observations at 27 sites in 7 European countries, this exploratory work is limited to only one site in Denmark.

3 Algorithm for validation of surrogate safety measures

The general algorithm is presented in Figure 2. It can be noted from the validation method list in the background section, that there are two parallel problems that require solutions: *i*) detection of the relevant situations in the video material and measurement of the surrogate safety indicators to be tested; *ii*) estimation of the expected number of accidents as the direct measure of safety at a studied site (the “ground truth” to compare with).

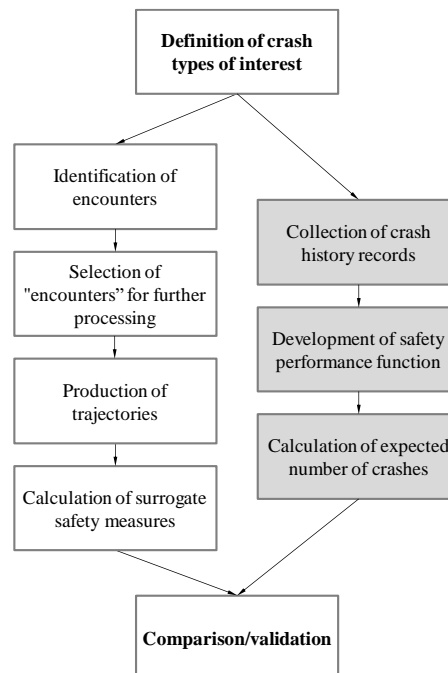


Figure 2. The general validation algorithm.

The crucial steps of the algorithm are further discussed in detail in the following section.

4 Case study - Denmark

4.1 Definition of crash types of interest

Even though it is desirable to have a universal surrogate safety measure applicable in all possible traffic situations, there are some quite strong evidences in literature suggesting that the relation between surrogate measures and crashes is not the same for different types of maneuvers and road users involved [33, 40, 41]. It was decided to start the analysis of the data disaggregated by crash type and take a decision on the eventual merging of the categories at a later stage. The drawback of the disaggregation is, of course, the risk of getting low numbers of crashes/surrogate events in each category: priority should therefore be given to the crash types that are relatively

frequent. Due to the specifics of the InDeV project, the following additional criteria were set: *i*) the crash must involve a motor vehicle and a vulnerable road user (pedestrian/ cyclist/ moped rider); *ii*) the crash must result in a fatality or serious injury; *iii*) the crash must occur at an intersection.

The latest version of the data structure for the European Union crash database CARE (CaDaS – Common Accident Data Set, [42]) contains a crash typology (a set of codes with corresponding sketches) that should have made finding the most frequent crash types fulfilling all the criteria a trivial task. However, this part of the CARE is still mostly empty as only two countries – German and Denmark – use similar systems at the national level and have the data available. Thus, the decision was made based on data from these two countries and complemented by the results of manually processed crash records from two large Swedish cities [43]. The finally selected crash types chosen are shown Figure 3. It is important to note here that these types are not the most frequent among all crashes, but only among those that fulfilled the stated criteria (the absolute leader of serious injuries were in fact single falls of pedestrians and cyclists - [43]).

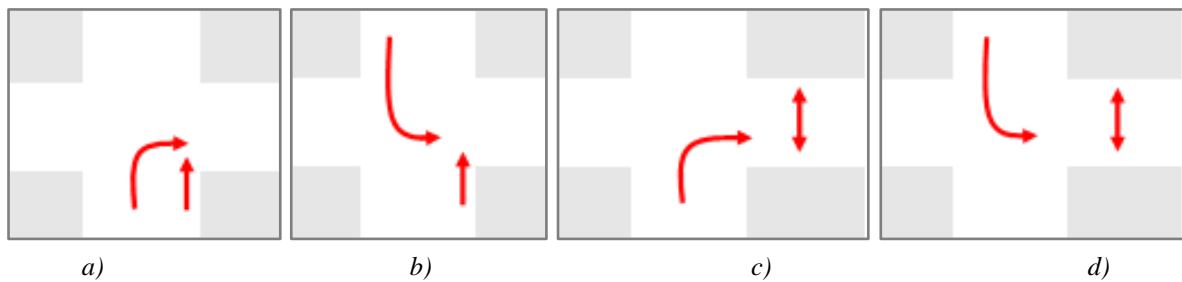


Figure 3. Most frequent crash types selected for further analysis: a, b) motor vehicle right/left - cyclist straight; c, d) motor vehicle right/left – pedestrian crossing the intersection approach.

4.2 Crash history data and safety performance functions

Quick examination of the crash history (period 2009-2013) from the selected sites revealed that such data is not sufficient since the total number of crashes is extremely low (once disaggregated, there is no accident in many categories). Therefore, it was decided to use a SPF based on a larger number of sites and having traffic flow as input. However, even the number of crashes from 50 similar locations is still extremely low (Table 1) and not sufficient for the construction of a completely new SPF.

Table 1. Disaggregated crash records from 50 Danish sites.

| | | | | |
|---|---|---|---|---|
| | | | | |
| Crash count from 50 similar sites (2009-2013) | 4 | 2 | 0 | 1 |

An alternative could be to update or adapt an already existing SPF to the given conditions. The systematic review of studies on safety-in-number effect by Elvik & Bjørnskau [44] provides a comprehensive list of existing models for crash prediction at intersections with pedestrian, bicycle and motor vehicle volumes as input. They note that the regression coefficients are very consistent among the reviewed studies. Thus, as a starting point, the following equations might be suitable [44]:

$$\begin{aligned} \text{Expected number of bicycle accidents} &= C * MV^{0.5} * CYCL^{0.43}, \\ \text{Expected number of pedestrian accidents} &= C * MV^{0.5} * PED^{0.51} \end{aligned}$$

where MV – annual daily motor vehicle traffic;

CYCL, PED – annual daily cyclist and pedestrian traffic

C – scaling parameter.

Methods for updating accident models can be divided into two categories: *i*) recalibration of the scaling parameter only and *ii*) re-calibration of all the SPF parameters [45]. Multiple methods have been suggested for both categories [45-47]. Regardless of method used, the amount of additional data used to recalibrate a SPF is very important. The number of locations depends on the mean number of observed crashes as well as the standard deviation of the observed crashes [48]. In this study we collected the accident data from 50 sites which is in line with many other studies dealing with model recalibration [44-47]. It must be noted, however, that methodological literature

recommends to have much more data, for example according to Shirazi et al. [48], for given crash frequency at the selected 50 intersections the recommended number of sites is at least 700.

4.3 Identification of the encounters

An encounter is a simultaneous arrival of two road users heading towards a common “conflict zone” that can be seen as an elementary event in traffic (or a unit of exposure - [49]) and has the potential to develop into a crash. The goal of this step is to find all the encounters that can be observed in video for further processing. After initial pre-viewing of the situations taking place, it becomes apparent that “simultaneous arrival” does not always result in a clear interaction, therefore a set of additional operational rules were set on how to treat groups of cyclists/ pedestrians, situations of a “protected passage”, etc.

To detect the encounters in video, we applied a watchdog software RUBA [50]. The basic functional unit of RUBA is a detector (a certain area of the image monitored for changes). Detection of presence, motion or motion in a certain direction is possible. By strategically placing the detectors and defining the rules for temporal relations between the detectors’ activations, it is possible to find the “simultaneous arrivals” of the two road users.

Since the number of encounters during the filmed three-week period is rather high, it is very valuable if at least some indication of their severity can be produced already at this stage so that the most severe ones can be given a priority in the analysis. Having this problem in mind, two strategies for placing the detectors in RUBA have been considered:

- The individual detectors are placed *in* the conflict zone (Figure 4a).
- The individual detectors are placed *before* the conflict zone (Figure 4b).



Figure 4. Location of the RUBA detectors car left – cyclist straight conflict: a) in the conflict zone; b) before the conflict zone.

The advantage of the first approach is that the time difference between the end of the activation of the first detector and start of the activation of the second detector is virtually equal to Post-Encroachment Time (PET). However, closer examination of the encounters and their severity ranking obtained by this method revealed two problems. First, when PET values get low (below 1 second), both road users are still visible in the “conflict zone” creating complex motion pattern which is often misinterpreted by RUBA, particularly affecting the start and end times of the activation of both detectors. Thus, for the situations expected to be most relevant for our purposes, the PET estimates are most unreliable (usually, overestimated by 2-3 seconds). The second problem is related to the property of the PET-indicator as such. In case one road user brakes heavily just before the “conflict zone” to avoid a collision, it might take several second until it finally starts moving again and activates the detector. It means, again, that a very severe situation will not be ranked as severe and potentially will not be analyzed at all if the selection of situation would be based on the criteria of low PET-value.

In the second case, the detectors are placed before the “conflict zone” and thus the simultaneous arrival is better reflected regardless to what evasive action is taken at a later stage. However, since the speeds of the road users are unknown (except for the fact that speeds are high enough to be considered a motion), it is not possible to separate situations with high speeds and collision course from situations when the car is “crawling” forward giving the way to the cyclist.

Without the ability to reliably rank the detected encounters, it was decided to select one day for which ALL the encounters are processed. This provides a relatively good picture of the frequency distributions for the events of “average” severity, but very uncertain estimates for the “severe end” of the distribution which would be based on very few situations. To compensate for that, the remaining detections from all other days are watched by an

observer who picks out only the situations expected to be severe. The instruction given to the observer is to include rather than exclude the situation when in doubt to make sure that none of the relevant events is missed.

Figure 5 shows the detectors used for the first studied intersection. All the detectors are based on the directed motion except for the pedestrian that is a simple presence detector (pedestrians are too small and “noisy” to allow for stable motion direction detection). It was also necessary to have an additional car detector in front of the pedestrian crossing since the turning cars often stop in the middle of the intersection when interacting with on-coming traffic or bicyclists.

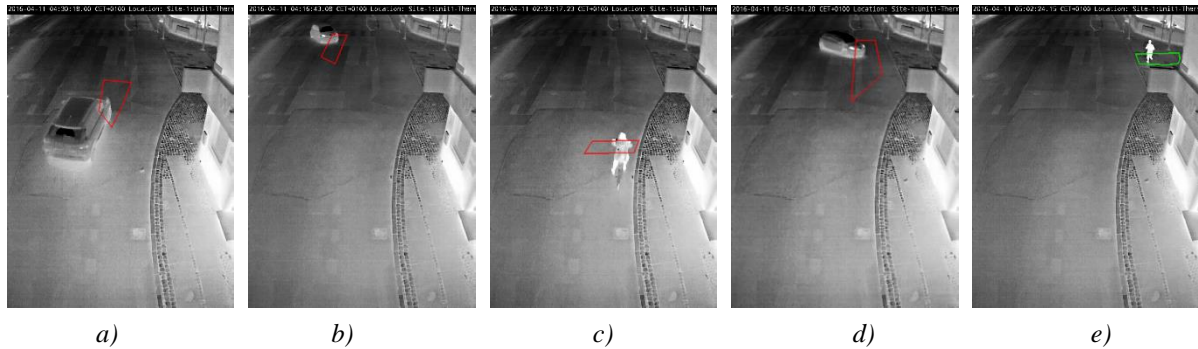


Figure 5. RUBA detectors used in the pilot study: a) right-turning car; b) left-turning car; c) cyclist; d) car in front of the pedestrian crossing; e) pedestrian.

Table 2 shows the results from the individual detections during five complete days compared to manual counts. It must be noted that the detectors were tuned so that the number of misses is minimal which results in a rather high amount of false alarms. It was not the purpose to use the detectors as a traffic counting instrument.

Table 2. Results of RUBA individual detectors compared to the manual counts (5 days).


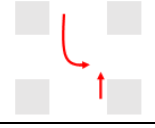

| | Manual counts (ground truth) | RUBA |
|---|---|--------------|
| Bicycles | 5885 | 7494 (+27%) |
| Pedestrians | 5595 | 9844 (+76%) |
| Cars left | 1576 | 3761 (+139%) |
| Cars right | 1783 | 3374 (+89%) |
| Cars in front of pedestrian crossing | 2610 | 5684 (+118%) |

The high number of false alarms for left-turning cars and the cars in front of the pedestrian crossing are explained by triggers from the straight going traffic (from left to right in the image). Since those are separated in time with the studied pedestrian/cyclist flows, these false alarms are easily removed at a later stage.

Table 3 shows the results of the encounter detector from RUBA based on temporal overlap (with a margin of 5 seconds) of the individual detector activations. For each encounter detection, an image containing two frames corresponding to the instances of activation of each individual detector was saved (like the ones shown in Figure 4). These turned out to be very efficient for visual checking of the results and removing situations when detectors were activated by birds, large vehicles, etc. The “automated detections” are compared to the manually selected encounters. Strictly speaking, after the visual check the automated detections are correct as they represent a simultaneous arrival of a car and a pedestrian/bicyclist. However, the additional rules set for treatment of the groups of bicyclists/pedestrians resulted in that the manual detections are much lower in amount. Only 5 situations detected manually were missed by the automated detector. In all cases, it was odd situations with bicyclist delayed for some reason in the middle of the intersection and thus having an encounter with a car arriving much later.

In total, the 24 hours of video was compressed to 1093 detection, which corresponds to approximately 4.5 hours (15 seconds per detection). Moreover, many of the detections overlap which means that the total “watching time” is even lower (estimated to 3.5 hours). That is an 85 % reduction in the amount of video to be manually checked.

Table 3. Results of RUBA encounter detections compared to the manual detections (1 day).

| | RUBA (auto-combined) | RUBA (after visual control) | Manual (ground truth) |
|---|--------------------------------|---------------------------------------|---------------------------------|
|  | 393 | 233 (-41%) | 65 |
|  | 496 | 373 (-25%) | 115 |
|  | 1146 | 487 (-58%) | 141 |

4.4 Trajectory production and calculation of safety indicators

For further analysis the semi-automated tool T-Analyst [51] is used. The software allows navigating through video frame by frame, marking each road user in the image and thus producing their trajectories (in world-co-ordinate system) and speed profiles. The screenshot of the program is shown in Figure 6. Processing of one encounter takes about 4-5 minutes.

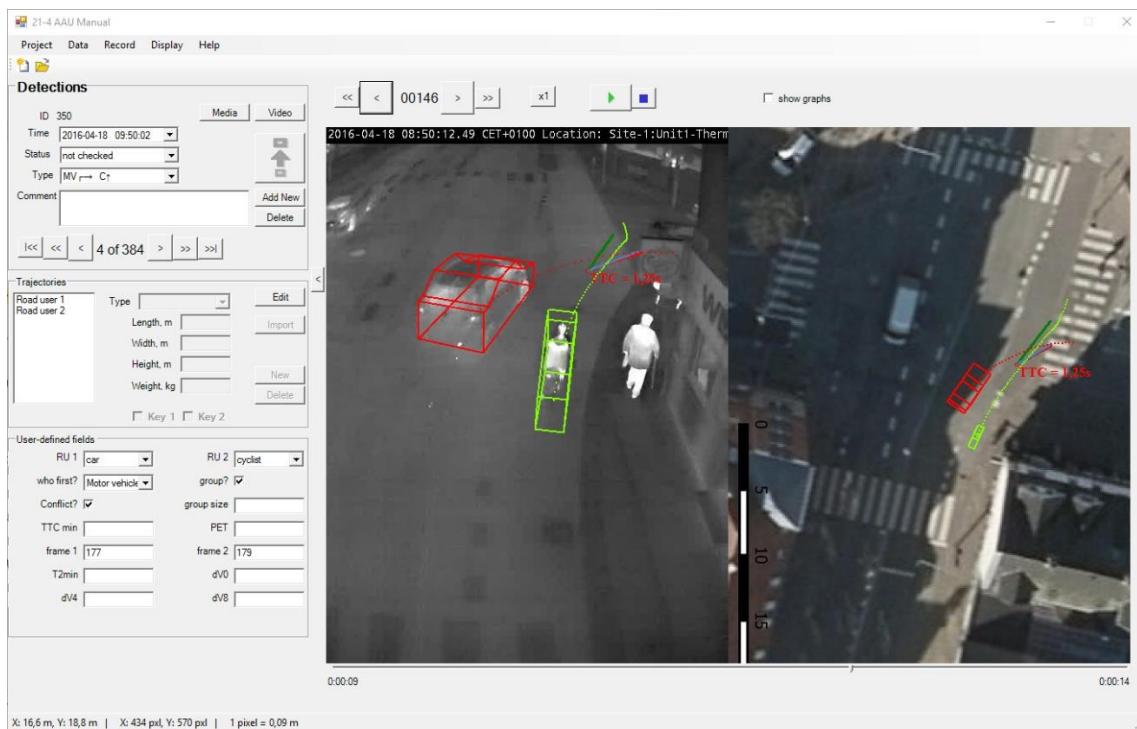


Figure 6. Screenshot of T-Analyst software [27].

Based on the trajectory and speed data, many surrogate safety indicators can be calculated. For this exploratory study, we selected the indicators that are already implemented in T-Analyst (but more indicators are planned to be tested):

- The minimum Time-to-Collision, TTC_{min} [52];
- Post-Encroachment Time, PET [11];
- T_2^{min} , time of arrival of the second road users to the potential collision point [53];
- Extended DeltaV₄ [6].

Not every indicator can be calculated for each situation (see Table 4), for example, TTC cannot be calculated for situations without a collision course.

The distribution of the calculated surrogate indicators are presented in Figure 7-Figure 10.

Table 4. Encounters and safety indicators that can be calculated.

| Total number of encounters | 65 | 115 | 71 | 70 |
|--|----|-----|----|----|
| Encounters with a calculable TTC_{min} value | 4 | 15 | 35 | 17 |
| Encounters with a calculable PET value | 65 | 115 | 71 | 70 |
| Encounters with a calculable T_2^{min} value | 65 | 112 | 65 | 70 |
| Encounters with a calculable Extended ΔV_4 value | 10 | 37 | 2 | 3 |

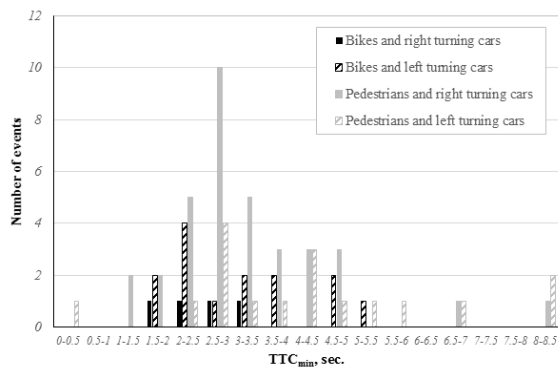


Figure 7. Distribution of TTC_{min} values.

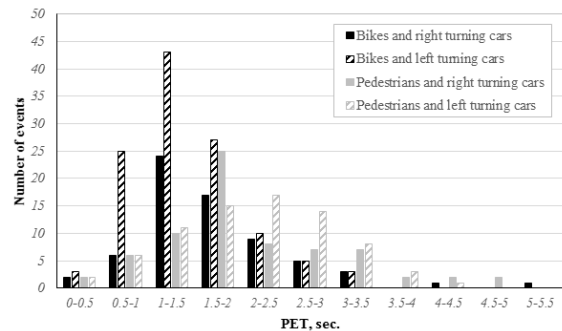


Figure 8. Distribution of PET values.

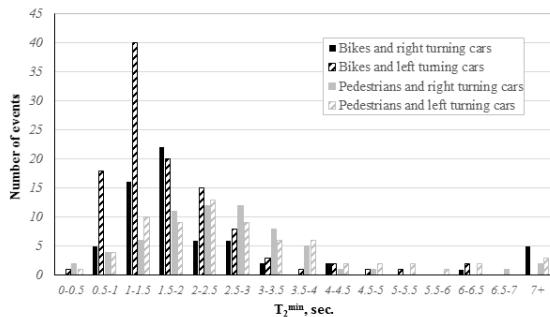


Figure 9. Distribution of T_2^{min} values.

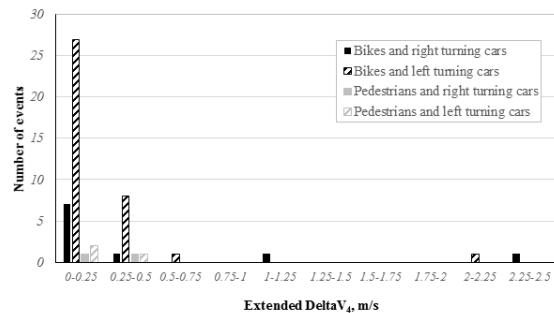


Figure 10. Distribution of Extended ΔV_4 values.

5 Discussion & conclusions

First of all, this work confirms that crash data at an individual site is too sparse to be used for any meaningful analysis. Even an attempt to develop a SPF using data from a large number of similar locations does not seem to be problem-free. The recommendations found in literature about 500-700 sites necessary to develop (or just update an existing) SPF are not practically feasible unless the main goal of the study is the development of the SPF itself and proper resources can be assigned to do such enormous data collection. From the perspective of the validation of surrogate safety measures, this creates a serious challenge: it is very difficult to prove that an indicator is an actual surrogate measure of safety if the “ground truth” has as much, if not more, uncertainty as surrogate-based predictions, as was found in some earlier validation studies [54, 55]. On the other hand, this is a very strong argument for further development and use of the surrogate safety methods as it appears that the crash data is not

about to improve in quality or increase in number in the future. To develop some other, “indirect” ways of validation of surrogate safety measures might be a future challenge for the researchers.

It is not possible to draw any conclusion on the validity of the tested surrogate safety indicators from only one site. However, some observations on the data characteristics can already be made. It is quite notable that TTC, the indicator most frequently used in surrogate safety studies, is calculable in very few cases. Our hypothesis is that, indeed, in many situations seeming to have a collision course the road users are in fact separated with a tiny time gap, which becomes apparent when TTC is calculated using correct dimensions of the road users and accurate and realistic trajectories. Therefore, T_2^{\min} appears to be much more relevant indicator as it can be calculated for situations with and without a collision course (T_2^{\min} equals TTC in case of a collision course). Another dimension of the problem is that in turning maneuvers for obvious reason the speed does not remain constant as the “classical definition” of TTC suggests [52]. Therefore, more advanced methods for future motion prediction might be necessary (as, for example, discussed in [56]).

Extended DeltaV is another promising indicator as it takes into account both the nearness to collision and the potential consequences of it. This is particularly relevant in case of vulnerable road users that are more prone to sustain injuries even in case of a minor collision. However, again, the number of situations that can be analyzed with this indicator is very limited.

The length of the observation periods that have been used for surrogate safety studies vary from few hours to several days [30]. In this project, we aim at analyzing three weeks of data per site which is to our knowledge the longest period ever used. However, the initial results indicate that for some conflict definitions (for example, commonly used $TTC_{\min} < 1.5$) the total number of serious situations might still be very low, indicating that possibly even longer observations should be used. This, however, might turn out to be a practical constrain that makes the method less attractive for practitioners. On the other hand, with sufficient automation of the data collection and processing, the additional cost for analyzing a few extra weeks might not be very large compared to the initial costs related to camera installations, calibration of the settings in the video processing tools, etc.

The tools used in this study have greatly improved the efficiency and accuracy of the video data processing compared to traditional work of human observers. However, some additional functions are still needed. For the watchdog RUBA it is crucial that some speed-related measure is introduced in the detectors so that high-speed situations (potentially more severe) can be distinguished from low-speed (and less severe) ones. The detection of the arrivals in its current version, though not perfect, makes the tool already quite efficient for removing irrelevant video parts. Applied on thermal video in this study, RUBA showed good performance in day and night conditions, which somewhat deteriorated in presence of rain and fog. Still, it appears that thermal data is more “stable” compared to traditional RGB video, which is considerably affected by light variations, shadows, glare on wet asphalt, etc. Closer examination of how the thermal data changes in precipitation conditions might help to stabilize the performance of RUBA during longer observation periods that would normally contain both sunny and rainy days.

The production of trajectories in T-Analyst is probably the most labor-intensive and mundane part of this study and automation of this task is highly desirable. On the other hand, in a “regular” surrogate safety study (not aiming at validating the method), the total number of selected events to process might not be very high and the additional time spent on the manual production of the trajectories might thus be balanced by the advantages of the higher accuracy obtained as well as the quality control of the produced data. Another benefits of careful examination of each detected event is a much better understanding of the interaction/crash process that otherwise might be lost when more automated methods are used.

Acknowledgment

This work is a part of the EU Horizon 2020 project InDeV (In-Depth understanding of accident causation for Vulnerable road users, grant agreement No. 635895).

References

1. Laureshyn, A. *Application of automated video analysis to road user behaviour*. Doctoral thesis. Transport & Roads, Department of Technology & Society, Faculty of Engineering, LTH, Lund University 2010. Bulletin 253.
2. Alsop, J. and J. Langley, *Under-Reporting of Motor Vehicle Traffic Crash Victims in New Zealand*. Accident Analysis & Prevention, 2001. **33** (3): p. 353-359. [http://dx.doi.org/10.1016/S0001-4575\(00\)00049-X](http://dx.doi.org/10.1016/S0001-4575(00)00049-X).
3. Amoros, E., J.-L. Martin and B. Laumon, *Under-Reporting of Road Crash Casualties in France*. Accident Analysis & Prevention, 2006. **38** (4): p. 627-635. <http://dx.doi.org/10.1016/j.aap.2005.11.006>.
4. Elvik, R., A. Høyte, T. Vaa and M. Sørensen, *The handbook of road safety measures*, 2009. Emerald Group Publishing Limited, Bingley, UK.
5. Tarko, A., G. Davis, N. Saunier, T. Sayed and S. Washington, *White paper surrogate measures of safety*, 2009.
6. Laureshyn, A., T. De Ceunynck, C. Karlsson, Å. Svensson and S. Daniels, *In search of the severity dimension of traffic events: Extended Delta-V as a traffic conflict indicator*. Accident Analysis & Prevention, 2017. **98**: p. 46-56. <http://dx.doi.org/10.1016/j.aap.2016.09.026>.

7. Johansson, R., *Vision Zero - implementing a policy for traffic safety*. Safety Science, 2009. **47**: p. 826–831. <http://dx.doi.org/10.1016/j.ssci.2008.10.023>.
8. Lareshyn, A., M. de Goede, N. Saunier and A. Fyhri, *Cross-comparison of three surrogate safety methods to diagnose cyclist safety problems at intersections in Norway*. Accident Analysis & Prevention, 2016 (in press). <http://dx.doi.org/10.1016/j.aap.2016.04.035>.
9. Ogawa, K., *An Analysis of Traffic Conflict Phenomenon of Bicycles Using Space Occupancy Index*. Journal of the Eastern Asia Society for Transportation Studies, 2007. **7**: p. 1820-1827.
10. Hayward, J.C. *Near misses as a measure of safety at urban intersections*. Master thesis. Department of Civil Engineering, The Pennsylvania State University 1971.
11. Allen, B.L., B.T. Shin and P.J. Cooper, *Analysis of traffic conflicts and collisions*. Transportation Research Record, 1978. **667**: p. 67-74.
12. Kraay, J.H., A.R.A. van der Horst and S. Oppe, *Manual conflict observation technique DOCTOR*. Foundation Road safety for all, The Netherlands, 2013.
13. Brown, G.R., *Traffic conflicts for road user safety studies*. Canadian Journal of Civil Engineering, 1994.
14. Bagdadi, O., *Estimation of the severity of safety critical events*. Accident Analysis & Prevention, 2013. **50**: p. 167-174. <http://dx.doi.org/10.1016/j.aap.2012.04.007>.
15. Alhajyaseen, W.K.M., *The Integration of Conflict Probability and Severity for the Safety Assessment of Intersections*. Arabian Journal for Science and Engineering, 2015. **40** (2): p. 421-430. <http://dx.doi.org/10.1007/s13369-014-1553-1>.
16. Zheng, L. and K. Ismail, *A generalized exponential link function to map a conflict indicator into severity index within safety continuum framework*. Accident Analysis & Prevention, 2017. **102**: p. 23-30. <http://dx.doi.org/10.1016/j.aap.2017.02.013>.
17. Hydén, C. *The development of a method for traffic safety evaluation: the Swedish traffic conflict technique*. Doctoral thesis. Department of Traffic Planning and Engineering, Lund University 1987.
18. Grayson, G.B., *The Malmö study: a calibration of traffic conflict techniques*. SWOV, 1984. R-84-12.
19. Lightburn, A. and C.I. Howarth, *A study of observer variability and reliability in the detection and grading of traffic conflicts*. Second international traffic conflicts technique workshop, 1979, 10-12 May 1979, Paris, France.
20. Malinovsky, Y., Y.-J. Wu and Y. Wang, *Video-based Vehicle Detection and Tracking Using Spatiotemporal Maps*. Transportation Research Record, 2009. **2121**: p. 81-89. <http://dx.doi.org/10.3141/2121-09>.
21. Saunier, N. and T. Sayed, *Automated analysis of road safety with video data*. Transportation Research Record, 2007. **2019**: p. 57-64. <http://dx.doi.org/10.3141/2019-08>.
22. Tarko, A.P., M.A. Romero, V.K. Bandaru, K.B. Ariyur and C.G. Lizarazo, *Feasibility of Tracking Vehicles at Intersections with a Low-end LiDAR*. 96th TRB Annual Meeting, 2017, 8-12 January 2017.
23. Ismail, K., T. Sayed and N. Saunier, *Automated safety analysis using video sensors: technology and case studies*. Canadian Multidisciplinary Road Safety Conference, 2010, Ontario, Canada.
24. Kastrinaki, V., M. Zervakis and K. Kalaitzakis, *A survey of video processing techniques for traffic applications*. Image and Vision Computing, 2003. **21**: p. 359-381.
25. Madsen, T.K.O., *RUBA - video analysis software for road user behaviour analysis*. 29th ICTCT workshop, 2016, 20-21 October, Lund, Sweden.
26. Suzuki, K. and H. Nakamura, *TrafficAnalyzer - the integrated video image processing system for traffic flow analysis*. ITS World Congress, 2006, 8-12 October, London, UK.
27. T-Analyst (2017) Software for semi-automated video processing. www.tft.lth.se/software Acc. 04 Apr 2017.
28. Gimm, K., *AIM Research Intersection – Infrastructure for research on interacting urban traffic*. 27th ICTCT Workshop, 2014, 16-17 October, Maribor, Slovenia.
29. Brinberg, D. and J.E. McGrath, *Validity and the research process*, 1985. SAGE Publications, Inc, Beverly Hills.
30. Lareshyn, A., C. Johnsson, T. De Ceunynck, Å. Svensson, M. de Goede, N. Saunier, P. Włodarek, A.R.A. van der Horst and S. Daniels, *Review of current study methods for VRU safety. Appendix 6 – Scoping review: surrogate measures of safety in site-based road traffic observations*. InDeV, Horizon 2020 project, 2016. Deliverable 2.1 – part 4.
31. Baker, W.T., *An evaluation of the traffic conflicts technique*. 51st Annual Meeting of the Highway Research Board, 1972.
32. Lord, D., *Analysis of Pedestrian Conflicts with Left-Turning Traffic*. Transportation Research Record, 1996. **1538**: p. 61-67. <http://dx.doi.org/10.3141/1538-08>.
33. Hydén, C., *A traffic conflict technique for determining risk*. University of Lund, Lund Institute of Technology, Department of Traffic Planning and Engineering, 1977. LUTVDG/(TVTT-3005)/1-95/(1977) & Bulletin 15 B.
34. Hauer, E. and P. Gärder, *Research into the validity of the traffic conflict technique*. Accident Analysis & Prevention, 1986. **18** (6): p. 471-481. [http://dx.doi.org/10.1016/0001-4575\(86\)90020-5](http://dx.doi.org/10.1016/0001-4575(86)90020-5).
35. El-Basyouny, K. and T. Sayed, *Safety performance functions using traffic conflicts*. Safety Science, 2013. **51** (1): p. 160-164. <http://dx.doi.org/10.1016/j.ssci.2012.04.015>.
36. Songchitruksa, P. and A. Tarko, *The extreme value theory approach to safety estimation*. Accident Analysis and Prevention, 2006. **38**: p. 811-822. <http://dx.doi.org/10.1016/j.aap.2006.02.003>.
37. Sacchi, E., T. Sayed and P. de Leur, *A comparison of collision-based and conflict-based safety evaluations: the case of right-turn smart channels*. Accident Analysis & Prevention, 2013. **59**: p. 260-266. <http://doi.org/10.1016/j.aap.2013.06.002>.
38. Elvik, R., *A negative binomial regression model of safety-in-numbers as evidenced in conflict data*. Institute of Transport Economics, Norwegian Centre for Transport Research, 2016. Working paper 50892.
39. InDeV (2015-2018) In-depth Understanding of Accident Causation for Vulnerable Road Users. EU Horizon 2020, Framework Programme for Research and Innovation, project No. 635895. www.indev-project.eu Acc. 2017-02-01.
40. Linderholm, L., *Vidareutveckling av konflikttekniken för riskbestämning i trafiken* (in Swedish) *Further development of conflict technique for in traffic risk assessment*. Lund University, Dept. of Traffic Planning and Engineering, Lund, Sweden, 1981. Rapport 7025.
41. Gärder, P., *Konfliktstudier i landsvägskorsningar* (in Swedish) *Conflict studies in rural intersections*. Institutionen för Trafikteknik, LTH, Lund, 1982. Bulletin 42.
42. CaDaS, *Common accident data set. Referece guide*, 2015. Version 3.4 - April 2015.
43. Björnberg, A. *Do we learn more from situation based accident coding? Case study of Malmö and Gothenburg*. Degree project. Transport and Roads, Department of Technology and Society, Faculty of Engineering, LTH, Lund University, Sweden 2016. Thesis 295.
44. Elvik, R. and T. Björnskaug, *Safety-in-numbers: A systematic review and meta-analysis of evidence*. Safety Science, 2017. **92**: p. 274–282. <http://dx.doi.org/10.1016/j.ssci.2015.07.017>.
45. Connors, R.D., M. Maher, A. Wood, L. Mountain and K. Ropkins, *Methodology for fitting and updating predictive accident models with trend*. Accident Analysis & Prevention, 2013. **56**: p. 82-94. <http://dx.doi.org/10.1016/j.aap.2013.03.009>.
46. Heydari, S., L.F. Miranda-Moreno, D. Lord and L. Fu, *Bayesian methodology to estimate and update safety performance functions under limited data conditions: A sensitivity analysis*. Accident Analysis & Prevention, 2014. **64**: p. 41-51. <http://dx.doi.org/10.1016/j.aap.2013.11.001>.

47. Wood, A.G., L.J. Mountain, R.D. Connors, M.J. Maher and K. Ropkins, *Updating outdated predictive accident models*. Accident Analysis & Prevention, 2013. **55**: p. 54-66. <http://dx.doi.org/10.1016/j.aap.2013.02.028>.
48. Shirazi, M., D. Lord and S.R. Geedipally, *Sample-size guidelines for recalibrating crash prediction models: Recommendations for the highway safety manual*. Accident Analysis & Prevention, 2016. **93**: p. 160-168. <http://dx.doi.org/10.1016/j.aap.2016.04.011>.
49. Elvik, R., *Some implications of an event-based definition of exposure to the risk of road accident*. Accident Analysis & Prevention, 2015. **76**: p. 15-24. <http://dx.doi.org/10.1016/j.aap.2014.12.011>.
50. Madsen, T.K.O., C. Bahnsen, M.B. Jensen, H. Lahrmann and T. Moeslund, *Watchdog System*. InDeV, Horizon 2020 project, 2016. Deliverable 4.1.
51. T-Analyst (2016) Software for semi-automated video processing. www.tft.lth.se/software Acc. 18 Feb 2016.
52. Hayward, J.C., *Near-miss determination through use of a scale of danger*. Highway Research Record, 1972. **384**: p. 22-34.
53. Laureshyn, A., Å. Svensson and C. Hydén, *Evaluation of traffic safety, based on micro-level behavioural data: theoretical framework and first implementation*. Accident Analysis & Prevention, 2010. **42**: p. 1637-1646. <http://dx.doi.org/10.1016/j.aap.2010.03.021>.
54. Migletz, D.J., W.D. Glauz and K.M. Bauer, *Relationships between traffic conflicts and accidents*. U.S. Department of Transportation, Federal Highway Administration, 1985. FHWA/RD-84/042.
55. Svensson, Å., *Vidareutveckling och validering av den svenska konflikttekniken* (in Swedish) *Further development and validation of the Swedish traffic conflict technique*. Lund University, Institute of Technology, Dept. of Traffic Planning & Engineering, 1992.
56. Mohamed, M.G. and N. Saunier, *The impact of motion prediction methods on surrogate safety analysis: A case study of left-turn and opposite-direction interactions at a signalized intersection in Montreal*. Journal of Transportation Safety & Security, 2016 (**online**): p. 1-23. <http://dx.doi.org/10.1080/19439962.2016.1255690>.