Annex C to FESTA Handbook

Version 5

Defining crash relevant events in NDS/FOT studies formulation

FOT-Net
(Field Operational Tests Networking and Methodology Promotion)

ABA	Aggregation Based Analysis
ACC	Adaptive Cruise Control
ACEA	European Automobile Manufacturers' Association
ADAS	Advanced Driver Assistance Systems
CAN	Controller Area Network
CRE	Crash Relevant Events
EBA	Event Based Analysis
EuroNCAP	European New Car Assessment Programme
FCW	Forward Collision Warning
FESTA	Field opErational teSt supporT Action
FOT	Field Operational Test
GMM	Gaussian Mixture Models
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSM	Global System for Mobile communications
HMI	Human Machine Interface
I2V	Infrastructure to Vehicle (see Vehicular communication systems)
ICT	Information Communication Technology
ISA	Intelligent Speed Adaptation
ITS	Intelligent Transportation System
LDW	Lane Departure Warning

NDD   Naturalistic Driving Data

NDS   Naturalistic Driving Study

NHTSA   National_Highway_Traffic_Safety_Administration

OEM   Original Equipment Manufacturer (in this document used as synonym
                of vehicle manufacturers)

RR   Range Rate

TTC   Time To Collision (relevant parameter in preventive safety applications
                – see e.g. Prevent)

V2I   Vehicle to Infrastructure (see Vehicular communication systems)

V2V   Vehicle to Vehicle (see Vehicular communication systems)

V2X   Vehicle to X (Any Station) (see Vehicular communication systems)

VMS   Variable Message Sign

WYLFIWYF   What-You-Look-For-Is-What-You-Find

Defining crash relevant events in NDS/FOT studies formulation

1 Introduction

Non-intrusive logging and analysis of real driver behaviour in real traffic comes in two flavours, i.e. Naturalistic Driving Studies (NDS) and Field Operational Tests (FOT). For NDS, the key goal is usually to gain an understanding of crash causation mechanisms and in particular, which driver behaviours are associated with increased crash risk. For Field Operational Tests (FOT), the key goal usually is to evaluate whether one or more (usually in-vehicle) safety technologies has a detectable and significant influence on crash risk (be it positive of negative).

In NDS/FOT studies, massive amounts of vehicle and video data are usually collected on a continuous basis. Once processed and uploaded, this generates an enormous amount of data for analysis in the database. While this data obviously can be sliced and sorted in very many ways and for many purposes, the focus of this report lies on how to approach the analysis in a way that leads to an understanding of how driver behaviours and/or vehicle systems influence crash risk. To contrast this with other types of analysis, one example would be using the data to identify natural acceleration and deceleration patterns that can be used for fuel economy tuning of vehicle engines.

For NDS/FOT data, there are two general analysis approaches that can be taken toward uncovering behaviours and/or system influences that drive changes in crash risk. The first can be called Aggregation Based Analysis (ABA). The basic principle of ABA is to identify trends and/or patterns in performance measures that have been aggregated over longer time segments. A typical example of such an analysis would be whether average time headway or average travel speed changes when drivers use Cruise Control. If there is a significant change, then that change might in a next stage be used to predict possible changes in crash/injury/fatality risk if the effects were to be extrapolated to the general driver population. However, in this report, ABA types of analysis will not be addressed. The reason for this is that such extrapolations are very difficult to justify, due to the fact that the connection between average measures and crash risk is hard to establish. This will be further discussed below. The other analysis approach is Event Based Analysis (EBA). The basic principle of EBA is to identify shorter driving segments (typically in the order of 5-10 seconds), during which the risk of crashing is judged to be higher compared to other driving in the data set, and then to analyse these events further. These events are often referred to as Crash Relevant Events (CRE), since their occurrence is thought to be indicative of actual crash risk in one way or another.

In NDS, analysis of CREs usually focuses on establishing why they occur from a driver perspective. In particular, the aim is to find out whether some types of driver behaviour are overrepresented in CREs as compared to baseline events, i.e. shorter driving segments where crash risk is judged not to be elevated. If particular behaviours can be identified as occurring disproportionally often during CREs compared to baseline events, then it is generally assumed that they can be viewed as contributing to the elevated crash risk, which makes them targets for countermeasure development. In FOTs, the focus when analysing CREs is somewhat different. One key question is whether the frequency of

various CRE types goes up or down as a function of Advanced Driver Assistance Systems (ADAS), i.e. if drivers experience critical situations less (or more) often when using ADAS. The other key question is whether driver responses during the CREs that occur is different when ADAS are being used. For example, does accelerator release come earlier when drivers are given a collision warning; do they brake harder when warned, etc.…). These types of analysis are not the same as looking for risk increasing behaviours, something which is important to keep in mind when setting up an FOT study.

2 Defining Crash Relevant Events

From the above, it follows that a key element to NDS/FOT success is defining CREs in a proper way. If the events selected for analysis are indeed crash relevant, then extrapolation to the general driver population is indeed both possible and credible.

However, while this is fine in theory, identifying CREs in NDS/FOT data is a bit more difficult in practice. To begin with, the simplest and most direct measure of crash risk, i.e. actual crashes are incredibly rare events. Even if hundreds of drivers are being continuously observed during several years of driving, the statistically low likelihood of having an actual crash means that the number of crashes in the final database will be smaller than required for statistical analysis, even if the database itself is huge.

It follows that surrogate events have to be used. These events have to have very particular properties, i.e. they need to be critical situations where there is no actual crash, but where the event still unfolds in such a way that its presence can be used as an indicator of crash risk.

The iceberg ratio metaphor is typically used here. For example, if one assumes that there are 100 incidents for every accident in a certain working place, then a measured reduction of incidents by 50% could be used to predict a corresponding 50% accident reduction in the future, even if there's been no accident yet. A key challenge to analysis and interpretation of NDS/FOT data is therefore how to couple non-crash events to crash causation mechanisms. In principle, if the link between crash causation and a CRE type is weak, then any interpretation of the CREs frequency or behavioural content is correspondingly weak, and vice versa.

Ideally, one would therefore only use CRE types that are known with certainty to be predictive of actual crash involvement, i.e. for which it is legitimate to infer that a change in their frequency corresponds to a (proportional) change in crash risk. Unfortunately, precise CRE definitions with clear cut, undisputable connection to crash involvement have yet to be fully established (had they done so, a simple list below would have concluded this report nicely).

To illustrate; while hard braking events might seem a plausible candidate, in the VTTI 100 car study (Dingus, Klauer et al. 2006) it was not possible to reliably identify critical driving events based on hard braking alone, since hard braking also occurred in many situations not associated with elevated crash risk.

2.1   Approaches to Crash Relevant Event Detection

This problem of identifying relevant CREs is not new. A lot of effort in many projects has gone into developing algorithms, filtering techniques etc. that allow for efficient yet relevant CRE selection. The aim here is not to list each such CRE in detail. Instead, the aim is to describe the approaches to data analysis that they represent on a higher, grouped level, and then go through some of the pros and cons that each CRE group, or analysis approach, faces.
Selecting among these approaches to CRE definition is a major decision point for a project. Projects must make an informed and conscious decision as to which of the below approaches will best fulfil their goals, and then set up CREs correspondingly. There are pros and cons to each approach, some of which may have large impact on the results.

2.2   Approach 1 - Driver response based identification

The first approach can be called the "driver response based" approach, and it builds on the general idea that CREs can be identified from the way drivers respond to them.

The most common version of this is to look for extreme vehicle kinematics. The basic assumption is that drivers prefer to travel in comfort and generally will not expose themselves to kinematically drastic events unless necessary. Thus abrupt velocity and direction changes in the vehicle (e.g. hard accelerations /decelerations and/or rapid steering) or abrupt movements of the driver (are thus considered to be out of the ordinary, indicating an unplanned and urgent response to an unexpected situation (of course, this might capture "play" with the vehicle as well).

There are two main ways to identify such drastic manoeuvres. Principally, one can either look for momentary or sustained breaches of defined thresholds. Looking for momentary breaches means identifying all instances where a particular threshold is exceeded, regardless of the excess duration. As an example, one could look for all instances of Brake Pressure Jerk (BPJ), above a certain value.

Just looking at the momentary threshold breaches might however capture many false positives. Kinematic signals can have high momentary peaks within their normal operation interval (e.g. acceleration spikes due to potholes in the road), and vehicle sensors are more often than not rather "noisy", i.e. signal variation does not correspond to real variations in the parameter. To remedy this (apart from signal filtering, which is more considered signal pre-processing before database upload and hence not covered here), an often used approach is to add a minimum time during which the threshold has to be breached.

For example, for deceleration levels above 0.8G, one can add a criterion that it needs to stay above 0.8G for more than 0.5 seconds for the event to count as a CRE. Of course, if the restrictions put in place are too strict, no CREs will be found in the data. Thus it is always necessary to strike a balance between removing false positives and keeping true positives.

A less common, but potentially promising approach is to use what can be called the startle response in the driver to find events. The general idea is that unexpected traffic situations

that also include a perceivable threat to the driver triggers a response in the form of an adrenalin rush and a general tensioning of the body. This bodily "jerk" may be used as a tell-tale that the driver did not expect the situation and that s/he perceives it as genuinely threatening, even though s/he may not brake or steer the vehicle in a particularly dramatic way.

2.3   Approach 2 - Function response based identification

If the study is of the FOT kind, i.e. designed to assess the impact of one or more active safety functions, then a very natural approach to CRE identification is to use the function itself to detect CREs. After all, that is what the function is designed to do.

For example, if an FOT is set up to assess the effects of Forward Collision Warning (FCW) on crash risk, the warnings issued by the collision warning system (even though they are not shown to the driver in the baseline phase) can be used as event identifiers.

The downside to this approach is of course that any CRE that occurs outside the function's detection capacity or that occurs when the system is turned off (in the treatment phase) will be missing from the analysis. Relying on the system signals only thus makes it impossible to estimate the frequency of CREs which the function in principle needs to detect but in practice cannot, though their prevention would enhance traffic safety.

Looking at the other side of the coin, the advantage of the approach is that it is not a worry how to factor in function availability and usage in the safety analysis. Since the function only can do something when it is turned on and does detect a threat, warnings/interventions can only occur when the function is capable of delivering them. Hence the true availability and usage rates are automatically represented in the data set.

2.4   Approach 3 - Driving context based identification

A third approach to CRE identification is to avoid relying on driver or function responses and instead make it driving context based. The underlying assumption of this approach is that too small margins equal elevated crash risk. In other words, there exist situations where the safety margins are inherently so small that the slightest mistake or variation could lead to a crash. Since mistakes do occur, it follows that crashes will also occur under these vehicle and/traffic environment configurations, and prevention of whatever it is that leads to these small margins thus will enhance traffic safety.

The definition of what constitutes too small margins can be either static or dynamic. An example of a static approach is the one used in the Road Departure Crash Warning System Field Operational Test (LeBlanc, Gordon et al. 2006). When evaluating the influence of a Lane Departure Warning system, the researchers used the lane marker as a static boundary, i.e. something that by definition should not be crossed unless the crossing is intentional.

Following this definition they looked for all instances of driving where the vehicle was travelling at or above the speed where the system would be available and where the vehicle inadvertently left the lane (defined as crossing the lane marker without having the turn signals activated).

An example of a dynamic version of the same approach is Najm et al. (Najm, Stearns et al. 2006), who developed a CRE algorithm that defined too small margins as being within a certain kinematic envelope. CREs were selected by identifying all situations where the vehicles fell below a combined Time-To-Collision (TTC) and Range Rate (RR) threshold in relation to a lead vehicle. The basic idea was that if a vehicle is closer than X to another vehicle and simultaneously closing in faster than Y, then there is a conflict regardless of whether an action is taken or not.

Four levels of conflict intensity were defined, depending on the precise TTC and RR thresholds. The values for Y as a function of X were selected based on empirical data from normal and deliberately delayed driver responses on a test track in lead vehicle following situations (Kiefer, Shulman et al. 2003). In other words, they essentially picked a demarcation line which says that if a vehicle is in this dynamic situation, the driver is by definition in trouble since most drivers would already have braked by now.

The advantage of this approach is that the actual conflict definition is highly objective, in the sense that it is based on vehicle kinematics only and does not depend on how the driver responds to it. Thus, while the setting of the boundaries for the kinematic envelope that defines the conflict might take some discussion, there will not be later disagreement as to whether the event was crash-relevant or not.

However, this also constitutes the weakness of the approach. Most importantly, active drivers who pursue a more aggressive driving style will be overrepresented in the CRE selection, since they will cross the envelope boundaries more often than drivers who prefer larger margins in general. Of course, if one assumes that aggressive driving is the main contributing factor to traffic accidents then this is OK. If one does not however, then there will be a selection bias in the CRE identification process that has to be dealt with somehow.

For FOT studies, this is actually quite feasible because CRE rates and dynamics with and without ADAS can be compared on an individual basis; thus taking out the effect of individual driving style. In a NDS-oriented study however, there will simply be more CREs for those driving with smaller margins, and some other means of dealing with that have to be identified.

2.5 Approach 4 - Driving history based identification

A fourth approach towards CRE detection is to look for unusual events in a driving history perspective, either on the individual or the group level. The idea is that unusual events in a person's or group's driving history are unusual precisely because the drivers try to avoid such events. Hence they can be said to represent situations which the driver is unwilling to enter, presumably because at least a certain portion of them would lead to potentially unsafe situations.

The advantage of this approach is that it will find the most unusual events that occurred during the study for each person or group (depending on the setup), and it is not unreasonable to assume that those rare events were ones which the drivers would prefer to avoid happening in the future. The corresponding disadvantage is of course that those events may be special for other reasons than being safety-critical, so even if drivers in a statistical sense try to avoid them, they may have little or no connection to traffic safety.

3   Coupling CREs and crash risk – the CRE causation question

As already discussed in the introduction, due to the very low risk of actual crashes occurring, NDS and FOT data sets typically contain only a few real crashes. Hence, the use of surrogate events like non-crash CREs usually is motivated by referring to some version of Heinrich's classical triangle from his pioneering work on industrial safety in the 1930's (Heinrich 1931). Often, they cite Heinrich's proposed 600 to 1 ratio between incidents and accidents as a reason why incident studies may be a faster way to learn about problems that need solving.

However, what is often less clear in published studies is an exact description of how they connect the CREs they use to real crashes. Questions like whether the causation mechanisms are the same for the studied incidents as the real accidents, and if so, which data types must be collected to verify that the "right" CREs have been found is usually missing.

In all four approaches to CRE definition outlined above, there are certain underlying assumptions regarding this coupling between the CREs and real crashes. For example, to go with a driver response based CRE identification, it has to be assumed that kinematically drastic driver responses are predictive of crash involvement, e.g. that hard decelerations are the "bottom" of an iceberg where lead vehicle crashes are the top.

If instead a contextual approach is used, it has to be assumed that small margins are predictive of crash involvement, i.e. that small margin situations are the items at the bottom of the iceberg, where "no margin" situations constitute the top (i.e. actual crashes). Or if a driving history approach is chosen, it has to be assumed that rarely occurring combinations of kinematic (and other) values are predictive of crash involvement.

In other words, the approaches represent very different views of the coupling between CREs and crash risk. Which one (if any) is more appropriate remains to be determined on a project by project basis, as they all have different consequences for how the project findings can be extrapolated to the rest of the driving population.

4   What do the identified CREs represent?

The approaches described above also have different implications in terms of what the selected CREs represent. In the driver response based approach, CREs are identified based on how each driver evaluates the situation. For example, while one driver may brake hard at a certain time-to-collision threshold, another driver might not brake at all when at that time-to-collision value. Hence a selection of CREs based on this approach will include the first event but not the second (since the driver did not respond, it is by definition not an event).

Driver response based CRE selections will therefore reflect the normal variability in any driver population in terms of driving style, risk perception and capacity to respond. It follows

that representative selection of drivers becomes a key issue when using a driver response centred CRE selection approach.

The function based CRE selection approach does present an interesting conundrum in terms of what the identified CREs really represent. Simply put, due to sensing and/or threat detection algorithm limitations, the function may not capture all the critical situations you think it should. Hence it would be preferable if you would could write and run your own threat detection algorithm on the collected data in order to avoid what a function limitation based CRE selection bias.

However, this is easier said than done. The developers of the function under FOT assessment are more often than not clever people who have worked on the function for several years. This means that the alternative threat detection algorithm you develop has to beat their algorithm, while still being based on the same sensor data. Unless you're very lucky or a genius, it might be unwise to rely on being able to do that.

Another option here is of course to add additional sensing capability before running the FOT, i.e. sensors which have much better performance than those used by the function being assessed, and then run a threat detection algorithm on that sensor data once collected. However, the additional equipment cost might be quite large, so you really have to believe there exists a class of important CREs outside the evaluated functions current detection capability. Moreover, it also must be necessary for the project to find them. For example, if you're tasked with defining the remaining safety potential in terms of avoiding rear end crashes given a particular FCW function, then this makes sense. Give this some thought before adding additional sensing to the project cost.

In a driving context approach, CREs are identified independently of how a particular driver responds to the situation. This means all drivers are equally covered by the conflict definition, independently of their capacity or willingness to respond. It also means that drivers with a more aggressive driving style will be overrepresented in terms of how many events they contribute to the list of CREs selected for analysis. They simply end up in small margin situations more often.

Now, if you assume that small margins in and of themselves are predictive of crash involvement, it does not matter that some drivers contribute with more CREs than others, since by this logic, these drivers do have a higher crash risk. On the other hand, if you're dubious about this assumption, then this approach might not be for you to begin with.

The advantage of a driving history based CRE selection approach is that one can tailor the CRE selection to each participating individual. Drivers all respond in some way when a potentially dangerous situation is sprung on them; but driver responses can be expected to vary even in complete surprise situations. For example, while 0.7G might be a normal deceleration level for one driver, another might never go above 0.65G regardless of how critical the situation is. Thus, if a 0.65G deceleration occurs only once in a person's full driving history, one might draw the conclusion that it was a special event for that driver and that it warrants further examination.

A prerequisite for the driving history based approach is that basic driver behaviour is fairly stable, i.e. that the distributions are not too "flat". For example, if you drive all over the lane

rather than keeping fairly well to the lane centre, the tail of your lane position distribution will include both the situations when you ended up there unintentionally and would have preferred to be elsewhere, and the situations where you just ended up there and it didn't matter. If that is the case, the search algorithm will find most events simply due to a lot of inherent variability in the data, not because the driving situations as such have special, crash relevant, properties.

As an example, a technique called Gaussian Mixture Models (GMM - for an example from another domain, see Reynolds, 1995) was tested in the euroFOT project. Gaussian distributions were fitted to a number of logged vehicle parameters such as brake pressure and pedal positions. Rare situations were then identified by looking for combinations of unusual values; i.e. where the values for several parameters were at the tail of their respective distributions simultaneously.

The GMM approach was found to work much better for truck drivers than for car drivers. For one, truck drivers drive more consistently due to spending so much more time behind the wheel. Furthermore, they also have a much smaller kinematic space to drift around in due to the size and mass of their vehicles. In other words, end points in their distributions therefore are more likely to represent truly unwanted values rather than random occurrences due to normal variability. Hence this approach might be more applicable for professional drivers who show less variability over time, rather than lay people who drive much less.

Obviously, any CREs identified will be more credible if they somehow link to underlying mechanisms related to the driver's control, e.g. over the vehicle, over positioning relative to other traffic participants, etc.

One recent approach that illustrates this is the work by Gordon et al (2011). They assumed that single vehicle crash risk depends on the driver's lateral control. The study assumed that the underlying mechanisms leading to such crashes are the same as those that create variations in normal driving—especially those involving "disturbed" lane-keeping control. The road-departure crash problem was formulated according to the following set of hypotheses:

- Single-vehicle road-departure crashes occur only under conditions of disturbed control
- Naturalistic Driving Data (NDD) contain measurable episodes of disturbed control
- Crash surrogates exist and are based on a combination of objective measures of disturbed control, highway geometric factors, and off-highway factors
- Crash surrogates can be related to actual crashes

The study compared lane deviation, LDW and instances of very short distance to the road edge as crash surrogates, and found the latter to be the best, with LDWs as the second best and variability within the lane to be the least successful predictor. In other words, in this study the empirical findings were predicted from theory and also followed the theory.

5 Removing false positives

Since there are no perfect CRE identifiers, any event search in a NDS/FOT database is likely to come up with a mix of true positives (i.e. actually relevant events) and false positives (i.e. events, while captured by the search algorithm, turn out not to be coupled to the proposed accident causation mechanism after all). An example would be a search algorithm that identifies hard braking events, but which cannot in and of itself distinguish if there is a lead vehicle in front or not.

Therefore, after a set of what could be called CRE candidates have been identified in any NDS/FOT data set, comes the very important step of weeding out the false positives from the true positives. There are essentially three ways to do this; manual (visual) inspection, filtering based on some logic, or a combination.

A seemingly straightforward way to remove false positives from among the event candidates' process is manual visual inspection. Researchers review each event based on the video recorded and other available data, and then decide whether the event is truly crash relevant or not.

One advantage of this approach is that variables that might be impossible to capture numerically can be weighed in, such as how startled or scared the driver seems to be. On the other hand, assessing a level of criticality from comparatively low resolution videos is often difficult; video is not reality after all.

From this also follows that the assessment will vary between researchers. Inter-rater reliability thus becomes an important factor. Given that a reliability round 80% is considered a good number in inter-rater reliability studies, one realizes that if one of every five events is potentially misclassified, the statistics used in the final analysis have to compensate for that (if possible).

Another way to go about it is by applying one or more conditional filters. This means that the initial group of CREs is screened based on whether the CREs fulfil certain additional criteria. In the lane departure example, in addition to having inadvertently left the lane, one might remove all CRE candidates where the driver did not respond to the situation by steering or braking within three seconds, on the assumption that if the driver fails to respond the situation is not critical at all (unless there is a subsequent crash of course).

Another example is the filtering applied by Benmimoun et al in euroFOT (Benmimoun, Fahrenkrog et al. 2011), where various combinations of speed dependent yaw and deceleration filters are used to identify lateral conflicts.

The search for a good CRE definition is often an iterative process, where the selection criteria are successively refined to suit the study purposes. As an example of such an iterative process, Fitch et al (Fitch, Rakha et al. 2008) used an iterative process where the initial pool of CRE candidates was screened using one new filter at a time. After each filtering, visual inspection of a randomly sampled portion of the remaining CRE candidates was carried out, and this process continued until the desired ratio between true and false positives was achieved.

5.1   Which way to go?

It is not obvious which of these approaches to filtering is the best way to go. Please also note that no formal restrictions apply, in the sense that if you select approach you must stick to that approach only. Some projects have used combinations of the above approaches to find events, and this is perfectly fine.

What is clear however is that different CRE definitions may lead to very different results, even when used on the same data. In the above study (Fitch, Rakha et al. 2008), it is described that another project (Hanowski, Blanco et al. 2008) used a kinematic threshold visual manual CRE candidate selection on the same data set as Fitch et al were using. The opportunity therefore arose to compare CRE selections between the two projects. Interestingly, while both projects found hundreds of what they judged to be relevant CREs, only 7 of the 596 conflicts found by Hanowski et al study were identified in the Fitch et al study.

This shows that if anything, the acronym WYLFIWYF (What-You-Look-For-Is-What-You-Find) coined by Erik Hollnagel (Hollnagel 2004)) holds for NDS/FOT analysis and CRE identification as well. It also shows that the most important requirement here is that you explicitly describe why you did choose the CRE definition you end up with in the end, i.e. how that particular CRE type is linked to the crash type(s) you are trying to understand and prevent.

6   The problem of Baseline selection

When analysing CREs, whether it is their relative frequency or some particular aspect of driver behaviour during the CREs, one always has to define a baseline, i.e. a comparison situation. This is one of the philosophically more challenging aspects of both NDS and FOT studies.

6.1   Naturalistic Driving Study baselines

In NDS studies, a key topic of interest is whether any particular driver behaviours are overly represented in CREs, i.e. whether they can be viewed as crash contributing factors. To find this out, one has to have a set of relevant comparison situations, i.e. a baseline, to compare the CREs to.

There are two principal ways to select baseline events (non-CRE events). One is to randomly sample a number of baseline events from all data that is not part of the CREs. The other is to tailor the baseline selection to the participants. For example, if a CRE occurs on a highway in the afternoon when it rains, then one tries to find a baseline event that also occurs on a highway in the afternoon when it rains. Naturally, this is easier for conditions that occur more often. Finding one or more matching baseline events for a CRE that occurred while commuting to work is easier than finding a match for a CRE that occurred on a vacation road trip.

More formally, these two approaches represent two types of experimental design (case-cohort vs. case-crossover designs) Rather than going into what can be expected given that either approach is used, the reader is referred to the in-depth comparison of the

strengths and weaknesses of each approach when used on the 100 car dataset (Guo and Hankey 2009).

6.2 FOT baselines

In a Field Operational Test, the baseline selection problem is slightly different compared to an NDS. The example of Adaptive Cruise Control (ACC) illustrates this point very well. It is relatively straightforward to say that the relative frequency of CREs that occur while ACC is in use constitutes the treatment data. However, what should that be compared to?

One way to resolve this is to focus on the adaptive part of the cruise control functionality. This means one would compare the relative CRE frequency when ACC is being used to the CRE frequency when regular cruise control is being used. Another way is to focus only on CREs that occurs in car following situations. Here, the baseline would be all CREs that occur when there is a lead vehicle present and ACC is not engaged.

More examples can easily be constructed, but these two suffice to illustrate that baseline selection inherently depends on the researchers perception of which the relevant accident causation mechanisms are.

As an example of how complex this may get, consider the following approach that finally was settled on and used in euroFOT:

The approach adopted in euroFOT was to exclude all treatment data in which ACC was OFF. This had the advantage that it ensures 100% usage in the treatment portion of the analysed data. However, it caused another problem: the driver selection bias. Since ACC usage is self-paced by drivers (i.e. the drivers themselves decide where and when to use it), the baseline data should principally be selected to include only driving where the drivers would have opted to use ACC, had it been available. Defining filters for selecting data according to this principle is very complicated and might be impossible, as it involves second guessing driver behaviour. Instead, a set of filters thought to approximate the perfect baseline data selection was used. These filters, which were applied to both baseline and treatment data, and their effect in equalizing the driving conditions, are described in the following:

6.2.1 *Car following filter*

In terms of safety, ACC is mainly targeted towards reducing the number and severity of rear-end crashes. Hence, only data from driving when a lead vehicle was present in front of the equipped vehicle was included.

6.2.2 *Posted speed filter*

Data in speed limits in which ACC was not used very often (usage below 25%) were discarded.

6.2.3 *Vehicle speed filter*

When approaching roundabouts and larger intersections, drivers normally brake. As this automatically disengages ACC and hence inevitably excludes this data from the treatment set (ACC on), it had to be removed from baseline as well. A simple way of tackling this issue is to set a limit on minimum vehicle speed. Drivers typically enter roundabouts and

larger intersection with speed equal to or below 50km/h, so by setting a minimum vehicle speed of 50km/h, most junction and roundabout driving was removed from both baseline and treatment periods.

6.2.4 *5 seconds wait-and-see filter*

As mentioned above, ACC disengages at braking. This means that harsh braking and critical time-gap events that happen right after ACC disengages would not be included in treatment if the ACC ON filter was strictly applied in treatment. To compensate for this, treatment data was selected to include five additional seconds each time ACC disengaged, to make sure these this type of events were coupled to ACC usage and not excluded from the treatment data.

7   Conclusions

Unfortunately, there is yet no formal CRE recipe available that allows just for doing and not thinking. Thus, as hopefully has become clear in the above discussion, setting up CREs and the corresponding baseline events requires thinking about crash causation mechanisms, driver selection principles, what actually constitutes risk-free driving, and many other things.

The aim of this chapter is to describe common approaches to CRE selection on a higher, grouped level, and list some of the pros and cons for each CRE group. The whole point of this exercise is to support and motivate you and your project to make an informed and conscious decision as to which approach to CRE selection best will fulfil your project goals, and then set up your CREs correspondingly.

However, let's not forget that progress is being made in many concurrent projects. Hopefully, we will see convergence toward common CRE definitions in the next few years. Also, while most work until now has focused on vehicle-to-vehicle conflicts, there are now also projects underway that will broaden the scope by looking at conflicts with vulnerable road users (pedestrians, bicyclists, etc.) as well as Powered Two-Wheelers. While each of these conflict objects represent new challenges when it comes to defining the actual crash causation mechanisms involved, as well as which surrogate events could be used as indicators of accident occurrence likelihood, they also drive a good portion of creative thinking on the subject, which is a good thing.

8 References

Benmimoun, M., F. Fahrenkrog, et al. (2011). Incident Detection Based on Vehicle Can-Data within the Large Scale Field Operational Test "euroFOT". 22nd Enhanced Safety of Vehicles Conference. Washington, D.C.

Dingus, T. A., S. G. Klauer, et al. (2006). The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment. Cambridge, U.S. Department of Transportation, John A. Volpe National Transportation Systems Centre.

Fitch, G. M., H. A. Rakha, et al. (2008). Safety Benefit Evaluation of a Forward Collision Warning System, US DOT, NHTSA, Report No DOT HS 810 910, Washington, D.C.

Tim J. Gordon; Lidia P. Kostyniuk; Paul E. Green; Michelle A. Barnes; Daniel Blower; Adam D. Blankespoor; Scott E. Bogard, Analysis of crash rates and surrogate events: Unified approach (2011) Transportation Research Record. 2011 ; (2237):1-9.

Guo, F. and J. Hankey (2009). Modeling 100-Car Safety Events: A Case-Based Approach for Analyzing Naturalistic Driving Data, The National Surface Transportation Safety Centre for Excellence.

Hanowski, R. J., M. Blanco, et al. (2008). The Drowsy Driver Warning System Field Operational Test: Data Collection Methods, Virginia Tech Transportation Institute.

Heinrich, H. W. (1931). Industrial accident prevention: a scientific approach. New York, McGraw-Hill.

Hollnagel, E. (2004). Barriers and accident prevention. Burlington, VT, Ashgate.

Kiefer, R., M. A. Shulman, et al. (2003). Forward collision warning requirements project: refining the CAMP crash alert timing approach by examining "last-second" braking and lane change maneuvers under various kinematic conditions.

LeBlanc, D., T. J. Gordon, et al. (2006). "Road departure crash warning system field operational test: methodology and results. Volume 1: technical report." 307 p.

Najm, W. G., M. D. Stearns, et al. (2006). Evaluation of an Automotive Rear-End Collision Avoidance System, Research and Innovative Technology Administration, Volpe National Transportation Systems Centre, National Highway Traffic Safety Administration: 425p.

Reynolds, D., Rose, R.C., (1995) Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Transactions on Speech and Audio Processing, Vol 3, No 1.