FIELD OPERATIONAL TESTS: CHALLENGES AND METHODS

Yvonne Barnard and Oliver Carsten
Institute for Transport Studies, University of Leeds
LS2 9JT LEEDS, UK
Y.Barnard@its.leeds.ac.uk, O.M.J.Carsten@its.leeds.ac.uk

ABSTRACT: Currently field operational tests are being performed in Europe in which vehicles are equipped with the systems under investigation. Participating drivers use the vehicles during their daily routines and large amounts of data are gathered by sensors capturing driver and vehicle behaviour, and driving context. A methodology for preparing and performing tests of in-vehicle systems has been developed within the European FESTA project. To support the development and execution of field operational tests, both at a European and national level, the European FOT-Net project organises knowledge transfer and networking. During seminars and workshops a large number of challenges and issues were identified concerning the definition of hypotheses, baselines and incidents, the challenges of testing multiple systems, dealing with large data-sets, and ethical problems. We will discuss some directions for solutions.

1 INTRODUCTION

The last years have seen a large increase in the development and introduction of advanced driver assistance systems (ADAS) and in-vehicle information systems (IVIS). Technology advances rapidly, but understanding the changes in driver behaviour resulting from using these systems requires research carried out in a rigorous and systematic way. Currently a substantial number of such studies are being carried out, mainly in Europe, the USA, Japan, and Australia. In Field Operational Tests (FOTs) vehicles are equipped with the systems under investigation and with recording equipment to collect data on driving behaviour and the environment. A FOT may be defined as “a study undertaken to evaluate a function, or functions, under normal operating conditions in environments typically encountered by the host vehicle(s) using quasi-experimental methods” [1]. Participating drivers use the vehicles during their daily routines, data are autonomously logged and the drivers do not receive special instructions about how and where to drive. Except for some specific occasions, there is no experimenter in the vehicle, and typically the study period extends over at least a number of weeks.

The European Commission has underlined the importance of performing FOTs, and several European projects have been (co-)financed in the 7th framework programme. To support the development and execution of FOTs, both at a European and national level, two support actions have been put in place: FESTA [2] and FOT-Net [3]. FESTA has developed a methodology for conducting FOTs, and FOT-Net supports the FOTs in progress by transferring and sharing knowledge, networking and training. In the following sections we will describe their work. During the interactions between members of FOT consortia, a large number of challenges and issues were identified. In Section 4 we will discuss some of these issues and directions for solutions.

2 FIELD OPERATIONAL TESTS

The European FOT-Net project has taken up the methodology developed in FESTA (see Section 3) and works on transferring the knowledge to FOT projects. The project is funded by the European Commission DG Information Society and Media under the 7th Framework Programme. It provides a platform for knowledge exchange and for strategic networking of existing and future national, European and international FOTs. FOT-Net collaborates with the FOT projects in order to support them, to build up a common knowledge base on how to perform FOTs, and to share the results and experiences. FOT-Net organises stakeholder forums and seminars, based on the methodology developed in FESTA. It has developed a database (in the form of a wiki encyclopaedia) with information on all known FOTs [4].

At the moment, in 2010, two large-scale pan-European FOTs are ongoing: euroFOT: European Large-Scale Field Operational Tests on In-Vehicle Systems [5], and TeleFOT: Field Operational Test of Aftermarket and Nomadic Devices [6]. At a national level some 30 FOT's have been identified which are currently running or have recently been completed. National FOT's are to be found in Belgium, Denmark, Finland, France, Germany, The Netherlands, Sweden, Spain and the UK. There is a large variety of systems being tested, to name a few: Intelligent Speed Adaptation, Adaptive Cruise Control, Lane Departure Warning, Forward Collision Warning, Eco-driving, Drowsiness Warning, eCall, information and navigation systems. Projects may be financed by the private sector (such as the car industry) or by national and European bodies, or by both. In the USA, Asia and Australia, FOT's are also being performed, with comparable systems.

3 THE FESTA METHODOLOGY

In the European FESTA project (Field opErational teSt supporT Action), a consortium of a large number of partners, both industrial and academic, a methodology was developed to conduct these studies. Using such a methodology helps to ensure a sound approach to conducting studies and obtaining reliable results, but also allows for data and results that may be compared between tests. In Figure 1 this methodology is summarised. A handbook was written in which the methodology is described in detail [1].The methodology consists of a process which systematically details the steps to be taken to set-up the test (the left-hand side of the V), the actual data acquisition (the bottom of the V), and the analysis of the data and evaluation and interpretation of the results (the right-hand side of the V).

The first part of the methodology to define the test follows a systematic research-oriented approach. After defining the functions and systems to be investigated (for example forward collision warning), the use cases have to be defined, specific events in which a system is expected to behave according to the specified function, for example car following. Use cases describe the boundary conditions under which a function is intended to be analysed. Next the research questions and related hypotheses need to be defined. Hypotheses should be statistically testable. The study is then designed in detail and performance indicators are selected. Performance indicators are quantitative or qualitative indicators, monitored at regular or irregular intervals, and can be compared to one or more criteria (for example acceleration). The next step determines which specific measures and sensors to use. When the whole test is defined, the actual data acquisition can take place. Data are stored in a database and analysed. Analysis leads to evaluation of whether the research questions have been answered and the hypotheses need to be rejected or not. The outcomes of the analysis should also answer questions about the functioning of the system, and intended and unintended behaviour. Finally the results should be scaled up to assess the socio-economic impact, answering the question what the impact would be if the system were fully deployed in a large proportion of vehicles (or in all vehicles). At several points ethical and legal questions should be addressed, such as the privacy of the participants, responsibilities in the case of system failure, etc.

For all these different steps in the FESTA V methodology, detailed recommendations, both of a theoretical and a practical nature, were developed.

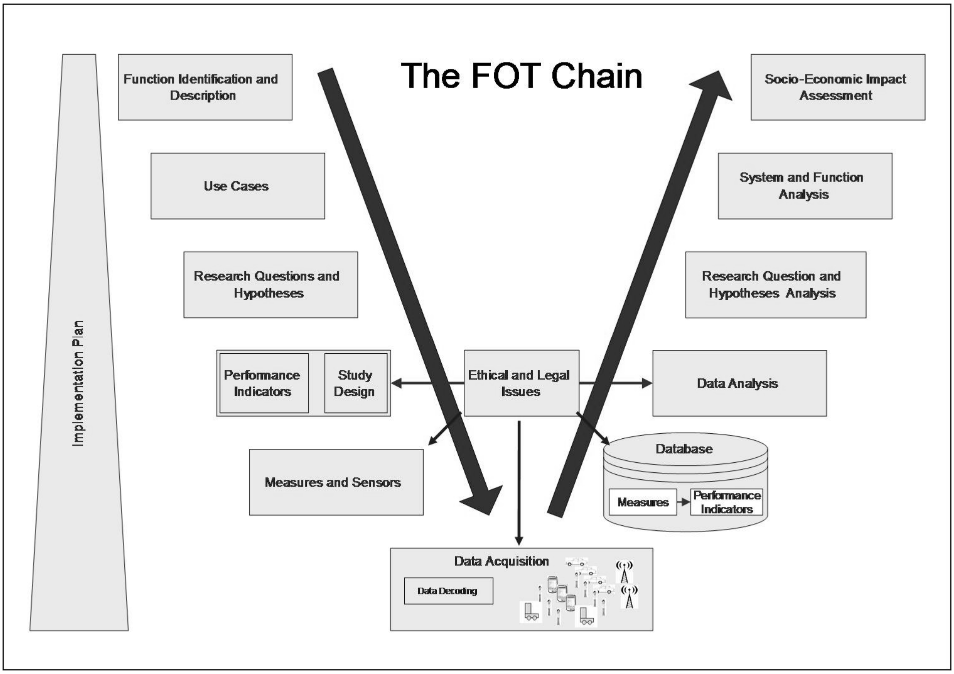

Fig. 1. the FESTA methodology

4 ISSUES AND CHALLENGES

During the seminars, stakeholder meetings and international workshops organised by FOT-Net, a large number of issues and challenges were identified, not all of them addressed by the FESTA methodology. We will discuss several of these in the following subsections.

4.1 Hypotheses

How to establish hypotheses on the impact of driving with systems, and how to choose the right ones from a large set of hypotheses?

Hypotheses are derived from research questions. FOTs are usually aiming to answer questions such as: Is it positive for the environment? Is it positive for safety? Is it negative for efficiency? Do drivers accept it? What is the market potential? What are the future costs and benefits? These general research questions need to be translated into more specific and testable hypotheses. There is no process that can assure that all the “correct” hypotheses are formulated. There may be different hypotheses per system, which contradict each other regarding the research question.

A combination of two complementary approaches has been tried out successfully several times: a top-down approach that considers six broad areas of system influence, and a bottom-up approach, applying use cases and situations to develop scenarios [7]. The top-down approach starts with the six areas of impact as defined by FESTA, based on [8] and is concerned with the direct effects of a system on the user’s driving and indirect (behavioural adaptation) effects of the system on the user as well as on non-users. Interaction between users and non-users (including vulnerable road users) may change. Accident consequences may also be influenced. Finally there are the effects of the combination with other systems. Starting from these impact areas, hypotheses may be formulated for different levels of the driving task: strategic (e.g. route choice), tactical (e.g. speed choice) and control aspects (e.g. steering behaviour) [9]. All three levels need to be considered. Consideration should be given to such mediating factors as driver state, experience, journey purpose, etc.

The bottom-up approach follows a systematic process as prescribed by FESTA, starting with the definition of the use case (e.g. car following) and situations in which the system would function (e.g. speed above 70 km/h and the Forward Collision Warning system is on). Next a set of detailed scenarios is defined, combining a use case and situations as well as an event that triggers a scenario (e.g. lead vehicle brakes sharply).

Finally hypotheses need to be prioritised. This may be a difficult process, as both approaches may give rise to a large number of hypotheses. Priorities are in the first place derived from relevance to the central research questions, and in the second place from practical and technical considerations.

Both approaches should be adopted, and iteration between them is necessary. In our experience workshops and discussions with interdisciplinary teams are very important in successfully formulating hypotheses as well as selecting the most useful and promising ones. Although formulating hypotheses may be hard work, in our experience it is also an interesting and inspiring activity when performed within a dedicated group.

4.2 Combinations of systems and functions

How to perform FOTs when using a combination of systems and functions that may interact with each other?

Some FOTs only address a single system with a single function, but others test driving with several systems and functions. Single system evaluation may become more difficult in the near future as more newly acquired vehicles will already be equipped with several systems. Defining hypotheses for combined systems and functions is more difficult, complex and time-consuming, and requires a clear idea about both the general research question and knowledge about the functioning of systems and possible interactions. Interactions may be on technical aspects, but may also be user-related. For example, the combination of a Forward Collision Warning system and a Traffic Information function could have a positive effect on safety and efficiency by increasing the headway kept, reducing the number and severity of accidents. It could also have negative effects by increasing drivers' workload and reaction time.

One way of addressing the challenge of defining hypotheses is to adopt a two-step approach: (1) identify and list hypotheses per function (isolated), and (2) then move to the combination of functions. Hypotheses about combined effects influence the experimental design. The number of comparison situations may have to increase when more than one system is available.

4.3 Baselines

How to establish a baseline for FOTs, with what situation should driving with a system be compared? And how to deal with cars equipped with several systems?

In order to be able to determine the effects of a system or function on the driving task, a baseline is needed during which drivers drive with systems off. This is not always easy, some systems cannot be turned off, or if people have bought a new car with a support system they may not be willing to turn it off. During a seminar, we discussed three different options concerning the problem of how to establish a baseline in FOTs, especially if more than one system or function is to be tested.

The first option is to have a control group with all systems off. Advantages identified are that this allows for a good and easy basis for comparison between groups and gives robust results. Some disadvantages are that it may be hard to find a representative control group, driving in similar conditions, and that it may be hard to keep this group motivated, and be sure that the systems are really off all the time.

Another option is to have no baseline at all. Some advantages are that the FOT becomes easier, quicker and cheaper to perform, but the disadvantage is that it is impossible to assess the impacts of the systems and their functions, so this is not a very useful option.

The third option is having several control groups with permutations of systems on and off. Advantages are that this option will yield more information, the different interactions may be measured, the effects of different systems can be separated, and (in)compatibilities identified. However, disadvantages are that the FOT will need more resources and becomes more complex. Some alternatives for these options are: comparing experimental groups of drivers with the different combinations, with one group that has the possibility to turn systems on and off, having a group with the permutations of systems off and on over time, using data from other studies to get information about some of the effects, and using data from naturalistic driving studies as a baseline.

4.4 Large data-sets

How to deal with the huge amount of data gathered in FOTs, how to analyse them?

Data in FOTs are automatically collected by equipment and sensors in the car such as the CAN-bus, radar, and video cameras. In addition, data may be collected by means of questionnaires, interviews, travel journals and observations. When participants drive for a considerable time a huge amount of data will be gathered. This poses all kinds of technical and practical challenges for transmitting, storing and analysing data. An important part of the FESTA handbook provides recommendations on these topics. FOT-Net organised a seminar dedicated to data management, and experts from different FOTs shared their experiences; their presentations have been made available [3]. Many of the questions from participants concerned issues of collecting, storing and using data. The first question is what data should be collected and stored. There are two extreme answers: collect only data needed to test the hypotheses, or collect as many data as possible. A “theory-only” approach may lead to throwing away or leaving out important and relevant data, and as FOTs are costly, it is often not possible to collect more data easily if needed. To “collect everything” may be impractical and costly, and at some point it will still be necessary to decide what data to use for analysis, because analysing all data may be impossible. The decision what data to collect is always a difficult one, and will to a large extent be driven by the resources available. The first priority of data collection is to collect data that are directly relevant for answering the research question. At the same time, it may be considered whether it is possible to collect some extra data without spending too much extra money. The FESTA methodology developed the Performance Indicator-Measures-Sensors matrix to support the selection of measures and sensors, providing support for making decisions.

FOT researchers have concerns about the possibility of loss of precious data. Lai, one of the experts with practical experience, gave the advice to put into place a monitoring protocol to minimise interruption to data collection [10]. Once the data are collected, there are two common approaches to minimise data loss: (1) multiple copies stored within the car – if one storage device fails, an identical set of data would still be available on a secondary storage device; (2) real time data transmission or regular access to the vehicle to obtain the data. And as there is always is the possibility of some disaster his advice is, based on experience: “Once the data are on the server, back up, back up, and back up again”

4.5 Incident definition

When investigating situations where the system plays a positive or negative role, how do we define incidents?

Many systems studied in FOTs have as their purpose to enhance safety and to prevent accidents. An example of a hypothesis from euroFOT is: "Using Forward Collision Warning, the number of forward crashes, near crashes and incidents will decrease". Because accidents do not happen very often, FOTs are interested in studying incidents which have the potential to lead to an accident. In euroFOT an incident is defined as: "an occurrence that could have resulted in a crash or near crash, if the circumstances would have been more adverse" [11].

It is necessary to be able to identify incidents in the continuous stream of data recorded in FOTs. It would, for example, save a lot of effort in analysing video-data if the analyst could focus on only those situations that pose a threat. Examples of data from sensors such as the CAN-bus, video and radar that could be related to an incident, are hard braking, rapid evasive manoeuvres, extraordinary close proximity to an object or pedestrian. However, this requires that triggers and thresholds are well defined, and that there are not too many false positives. For example, not every hard braking action is related to some possible dangerous situation. How to define and detect events of rather low severity is a question on which still more research is needed. Not all risky situations are characterised by discrete events. More work is also needed on developing continuous event severity scales and on mechanisms to detect them.

If it is too hard and costly to identify and categorise all incidents in the huge data set coming out of FOTs manually, data coming from sensors need to be processed and combined automatically. Advanced data mining techniques and statistical approaches may be needed to be able to do this. Such techniques are becoming available but are generally not yet deployed in FOTs. Hjälmdahl [12] concludes on the basis of experiences in the Swedish SeMiFOT that triggers need to be refined because of too many false positives and lack of information about missed positives. Inter-rater reliability needs to be improved, maybe through better training. Trigger parameters need to be optimised and adjusted for individual variations. New ways of identifying and validating incidents, especially low severity ones, are needed. An example is to look at the visual behaviour of drivers to predict incidents.

4.6 Ethical issues

How to deal with ethical and legal issues?

Because FOTs are a relatively new method of investigating driver behaviour, and tests are being conducted with ordinary drivers with everyday driving for a relatively long period, ethical and legal questions arise. A FOT is not the same as an experiment on the road in which an experimenter is sitting next to the driver. Special difficulties may arise when drivers cross national borders while driving an equipped car. Regulations and laws may differ between countries. For cross-border studies it may be possible to arrange dispensations for national regulations. There may also be some practical issues in this area, such as equipped cars with sensors may not be allowed to enter private sites or security sensitive sites (airports, military sites, etc.) for security reasons; even with the equipment turned off.

In FESTA special attention was given to ethical and legal issues, especially concerning participant recruitment, participant agreement, data protection, risk assessment, system safety, approval for on-road use, insurance, video data collection, and ethical approval [1; 13]. An important part of the issues that came up during discussions within FESTA and FOT-Net has to do with responsibility. For example, who is responsible if the participant has an accident or is fined by the police, or if the system does not function properly? Some of these questions have to do with the fact that the systems studied are quite new, for example, what if the system can override the decision of the driver (e.g. by braking), what are the responsibilities of the driver and the car manufacturer? Or how do you deal with testing a system that is not (yet) entirely mature or ergonomically optimal?

Many of these problems can be solved by making clear arrangements before the FOT starts, by performing a risk assessment, and by informing the drivers about these issues and having them sign a letter of agreement. Informing drivers is good, but has to be done carefully in order not to influence their driving behaviour, so mostly information needs to be given about the context of the test. For example, the Dutch FOT "The Assisted Driver" organised an information meeting at the start of the project. Participants were happy to be informed about the full context, and to be able to explain their participation to family and friends [14].

Other issues concern privacy. A lot of data are gathered from people that may be quite sensitive to their privacy, such as their destination of driving. Especially recording data near the participants' houses may be very sensitive. However, not recording them may lead to extensive loss of potentially interesting data. Most researchers agree that data should not be given to outside parties in the case of accidents, other than by court-order, and certainly not to insurance companies. Next to assuring that data are anonymous, well protected and not accessible for third parties, some challenges remain. For example: how to deal with video data of other road-users or passengers? These people do not participate and did not give permission.

An option for dealing with privacy and data ownership is to provide participants with the possibility to delete information that is still in the car and not yet uploaded. This may be of interest when the driver is involved in an accident or non-legal action. However, this option is highly disputed, because this may lead to loss of very interesting data, and for practical data-logging and data-transfer reasons.

5 CONCLUSIONS

FOTs are a relatively new way of performing research on how drivers behave when being supported by all kinds of systems within the car or on the road-side. Advanced equipment (for recording, transmitting and storing data) is rapidly becoming available and affordable, but this means that there is little experience with how to use them. Setting up and conducting FOTs is complex and requires a wide range of expertise. All this means that there are many open questions which do not have ready answers based on solid experience. There are two ways in which challenges may be addressed.

The first is by taking a systematic approach, following a methodology that decomposes the process of setting-up and conducting FOTs into well-structured steps, ensuring that important aspects are not neglected. FESTA provides such a methodology, but as it is the first to do so, not all the issues are addressed, and it is still in the process of practical try-out. There are gaps and issues to be explored further, some of which have been discussed in this paper, like defining incidents and dealing with multiple systems. In seminars we have conducted several workshops and exercises using a systematic approach. One of the most important experiences is that it is necessary to bring together a group of people from different disciplines, and to allow for open but structured discussions.

The second solution lies in the sharing of experiences, and building a knowledge base that is accessible for everyone who wants to conduct a FOT. Only if a body of knowledge both in the form of stored lessons learned and best-practices is built up and maintained will it be possible to develop further methodologies and recommendations. This is not only needed for methodological and scientific reasons. FOTs are still quite expensive to conduct and following best practices means avoiding having to re-invent everything from scratch, saving resources and becoming more efficient. Building on experiences also means having a network of researchers, technicians and other people involved in FOTs who can be asked questions and can be asked to become involved in new FOTs. FOT-Net contributes to this by organising meetings, providing information via its website and newsletter, maintaining a wiki and a forum, and above all by bringing people into contact with each other.

A final challenge is the re-use of data. As large amounts of data are gathered, and often only a part of them are analysed, many stakeholders and researchers are interested in using data from other FOTs to find answers to their specific questions. Gathering data in a FOT is costly, so if it is possible to share data and doing different kinds of analysis on them looking at different aspects, much may be learned about driver behaviour for relatively low costs. Re-use of data may sound easy but in practice raises many questions, such as: who owns the data, who is allowed access, how can data be stored in a format that facilitates re-use by others, who maintains and stores the data and makes them available to others. However, the lively discussion around these questions indicates that there is an important interest and willingness to share data within the FOT community.

6 REFERENCES

[1] FESTA Project: 'FESTA Handbook Version 2. Deliverable D6.4 of Field operational teSt supporT Action', 2008. Available at http://www.its.leeds.ac.uk/festa/

[2] FESTA: http://www.its.leeds.ac.uk/festa/index.php

[3] FOT-Net: http://www.fot-net.eu/

[4] FOT-Net Wiki: http://wiki.fot-net.eu/

[5] euroFOT: http://www.eurofot-ip.eu/

[6] TeleFOT: http://www.telefot.eu/

[7] Carsten, O., and Barnard, Y.: 'Preparing field operational tests for driver support systems: a research oriented approach', in D. de Waard, A. Axelsson, M. Berglund, B. Peters, and C. Weikert (Eds.): Human Factors: A system view of human, technology and organisation. (Shaker Publishing, Maastricht, the Netherlands, to appear in 2010)

[8] Draskóczy, M., Carsten, O., and Kulmala, R.: 'Road safety guidelines. Deliverable B5.2 of CODE project (TR1103)' (Atkins Wootton Jeffreys, Birmingham, UK. 1998)

[9] Michon, J.A.: 'A critical review of driver behaviour models', in L. Evans, L., and R.G. Schwing (Eds.): 'Human Behavior and Traffic Safety' (Plenum Press, New York, 1985), pp. 485-520

[10] Lai, F.: 'Data management: from car to database to analysis process', presentation at the FOT-Net seminar on data gathering and handling, 13 May 2009, Munich, Germany. Available at: http://www.fot-net.com/download/seminars/extraseminar/FOT-Net%20data%20seminar%20Lai.pdf

[11] Benmimoun, M.: 'Incident definition and usage within the euroFOT project', paper presented at the International Co-operation on Theories and Concepts in Traffic Safety (ICTCT) Conference, 22-23 October 2009, Leeds, UK. Available at: http://www.fot-net.eu/download/ICTCTconference/hjalmdahl_ictct_2009.pdf

[12] Hjälmdahl, M.: 'Incidents in SeMiFOT', paper presented at the International Co-operation on Theories and Concepts in Traffic Safety (ICTCT) Conference, 22-23 October 2009, Leeds, UK. Available at:

http://www.fot-net.eu/download/ICTCTconference/hjalmdahl_ictct_2009.pdf

[13] FESTA Project: 'FOT requirements, legal aspects, planning and development'. Deliverable D6.3 of Field operational teSt supporT Action. Available at: http://www.its.leeds.ac.uk/festa/downloads/FESTA_D6%203_Requirements_Legal_Ethical.pdf

[14] Alkim, T., Bootsma, G., and Looman, P.: 'The Assisted Driver: Systems that support driving' (Ministry of Transport, Public Works and Water Management, Delft, The Netherlands, 2007)