

Preparing field operational tests for driver support systems: a research-oriented approach

*Oliver Carsten & Yvonne Barnard
University of Leeds
Leeds, UK*

Abstract

A methodology for preparing and performing field operational tests (FOTs) of new-vehicle systems was developed by the European FESTA project. Currently several large FOTs are being performed in Europe and North America and Australia. With the growth of in-vehicle electronics and the advances in sensor technology, FOTs can collect huge amounts of data on driver and vehicle behaviour, as well as on driving context. Just collecting large quantities of data may not answer all the questions stakeholders have about the impacts of system use, and the data may also not lead to increased understanding. FESTA advocated a structured approach, starting with identifying the functions to be tested, followed by the definition of use cases and the formulation of research questions and hypotheses. All of this should precede definition of the experimental design of the study. However, in practice it turns out that it is not easy or straightforward to identify these questions and to formulate sensible hypotheses that can be tested statistically. To support the process of hypotheses generation, a systematic procedure was developed in FESTA: top-down, considering six areas of system influence, and bottom-up, developing scenarios. It is recommended that a multi-disciplinary team works jointly and iteratively with these two approaches. The practicality of investigating these hypotheses is subsequently checked by selecting appropriate performance indicators which can be used to confirm or refute them. The method has been successfully tried out in several workshops (including in the FOT-Net support action). The paper describes the FESTA methodology, the method for generating hypotheses and the experiences with the method. One issue discussed is the even more complex process of identifying hypotheses for a combination of driver support systems.

Introduction

Field Operational Tests (FOTs) are large-scale evaluations of new driver assistance systems, which aim to provide confirmation or refutation of the benefits of these systems in terms of safety, the environment and traffic operations, as well as of such user-related aspects as usage rates, driving comfort, acceptance and willingness to pay. They often culminate in a cost benefit analysis. The FESTA Handbook (FESTA Project, 2008) defines an FOT as “a study undertaken to evaluate a function, or

In D. de Waard, A. Axelsson, M. Berglund, B. Peters, and C. Weikert (Eds.) (2010). *Human Factors: A system view of human, technology and organisation* (pp. 1 - 6). Maastricht, the Netherlands: Shaker Publishing.

functions, under normal operating conditions in environments typically encountered by the host vehicle(s) using quasi-experimental methods”.

Laboratory experiments in the driver behaviour and driver human factors areas, such as those conducted on driving simulators, are normally designed around an *a priori* set of hypotheses, so that the participant selection, experimental design, experimental situations and events and the parameters collected are all tuned to the hypotheses being tested. By contrast in an FOT, the initial research questions are much more broad –e.g. is there a positive effect of the system on safety?– and the experimental design may be determined far more by what is possible or feasible. For example, if evaluating systems that are already on the market, an FOT may be forced to use as participants those who have purchased or are using vehicles with the relevant equipment. Similarly an FOT may for reasons of practicality not be able to use a fully counterbalanced design for comparing driving without system or function to driving with system or function. Indeed, FOTs typically compare a before period when the system is not operational to an after period when it is operational.

On the other hand FOTs are not purely *post hoc* studies, in which hypotheses are created after the fact and data is explored to confirm or refute those hypotheses. They are planned investigations in which considerable effort and cost is typically expended on the procurement and installation of data acquisition systems, the specification of the data to be acquired and the actual collection of data. As a result it is clearly sensible for an FOT to refine and prioritise its hypotheses in advance so as to ensure that the data collection is fit for purpose.

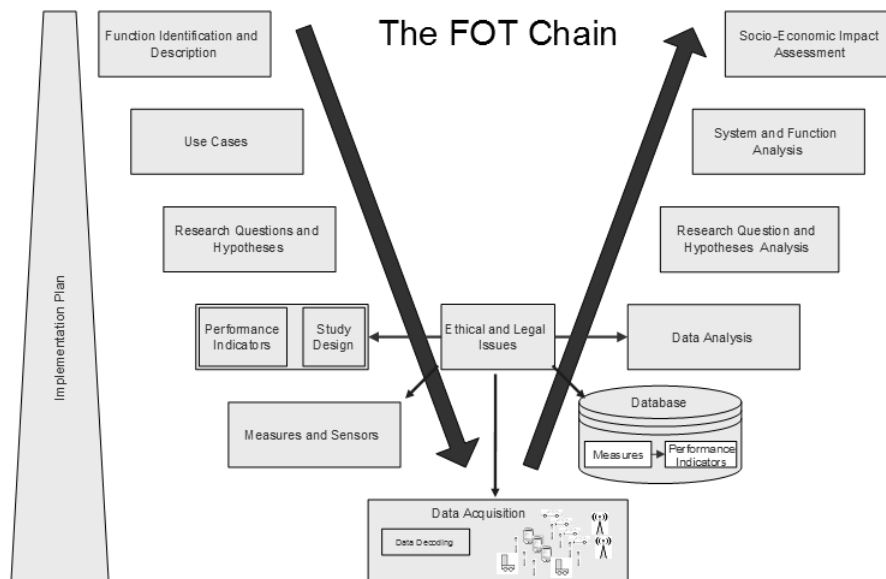


Figure 1. The steps that typically have to be considered when conducting an FOT. The large arrows indicate the time line (source: FESTA, 2008).

The FESTA project created a V diagram to summarise the recommended steps in preparing and executing an FOT. This diagram is shown In Figure 1. It can be seen that FESTA clearly recommended the specification of *a priori* hypotheses before the detailed specification of the study design and data collection procedures.

Identifying hypotheses in a structured manner

The creation of hypotheses about system effects on driving, on traffic and on user acceptance is hardly a matter of objective science. Different researchers will have different sets of interests and different opinions about likely system impacts. In some cases, they may be able to cite prior evidence to support their views; in other cases, those views may be based on mere hunches. Thus hypothesis formulation is arguably more of an art than a science, although it becomes more scientific as research on a particular system proceeds and previously proposed hypotheses are confirmed or rejected.

In addition it is not particularly difficult for a group of researchers to generate dozens if not hundreds of hypotheses. The potential for such an explosion of hypotheses can be illustrated by an example from naturalistic driving studies, which are applying similar data collection methods to those used in FOTs to the study of “normal” driving, i.e. driving without a driver assistance system. Task S05 of the U.S. SHRP2 research programme is concerned with designing the large-scale real-world study. That task enumerated hundreds of open-ended research questions before it even proceeded to the hypothesis stage (VTTI, 2007).

Definition of a hypothesis

FESTA distinguished between more general and open *research questions* and more specific *hypotheses*. These terms and many others were defined in the project glossary. This glossary has been further developed in the current EuroFOT project and the EuroFOT definitions are used here. The definition of a research question is “a general question to be answered by compiling and testing related specific hypotheses”. An example would be: “Does having a Forward Collision Warning system improve safety in driving?”

A hypothesis is defined as “a specific statement linking a cause to an effect and based on a mechanism linking the two. It is applied to one or more functions and can be tested with statistical means by analysing specific performance indicators in specific scenarios. A hypothesis is expected to predict the direction of the expected change.” The term “function” is used because a particular system may have a number of distinct functions — for example, one system could provide both Adaptive Cruise Control and a Forward Collision Warning. It is also the case that one function can be provided by different systems. An example of a hypothesis might be: “Forward Collision Warning will have the direct effect of an increase in minimum Time to Collision (TTC).”

It is important to note that the term “hypothesis” as used here differs from that often employed in statistical manuals. These tend to discuss hypothesis testing as being

about relationships that do not necessarily imply causation: a hypothesis that is confirmed (or a null hypothesis that is rejected at a certain level of significance) does not indicate causation but merely a relationship.

However, it makes little sense to apply this principle to the generation of hypotheses *a priori*. Here hypotheses can only be generated in a scientific manner by drawing on theory to support the linkage between cause and effect. Otherwise hypotheses become little more than wild guesses. Elvik (2008), discussing road safety modelling, distinguishes between non-causal statistical relationships and causal relationships, proposing that the distinguishing characteristics of causal relationships are a mechanism generating the statistical association between cause and effect and a relationship that is theoretically plausible or at least not inconsistent with well-established scientific theory. This definition is very close to that of EuroFOT.

There is also some approach to hypotheses in the wider social science literature. Thus Markus (2008) states: “To fully specify their hypotheses, researchers need to do more than just specify [a structured equation] model and plug in their substantive variables. They need to interpret the effect coefficients by specifying the type of causation that links the variables.”

Complementary approaches to generating hypotheses

In various workshops, initially in FESTA and subsequently in FOT-Net, it has been found that two complementary approaches provide a useful and helpful structure for generating hypotheses. It is possible to iterate between the two approaches.

The first approach is a *top-down* one, based on a structure outlined by Draskóczy et al. (1998). That structure provided ten potential areas of safety impact for a system or function. In FESTA, with a view to eliminating some potential overlaps, those were reduced to six areas of potential impact:

1. Direct effects of a system on the users and driving
2. Indirect (behavioural adaptation) effects of the system on the user
3. Indirect (behavioural adaptation) effects of the system on the non-user (imitating effect)
4. Modification of interaction between users and non-users (including vulnerable road users)
5. Modifying accident consequences (e.g. by improving rescue, etc.)
6. Effects of combination with other systems

FESTA also proposed some further considerations to apply in deliberating on these six areas. First of all the three levels of the driving task — strategic, tactical and control (Michon, 1985) — should be considered. For the strategic level, the potential to modify mode choice, route choice and exposure (frequency and/or length of travel) should be considered. At the tactical and control levels, the potential of modifying speed choice should be addressed. And finally, mediating factors such as experience and attitudes should be considered, since they might, for example, indicate particular effects for certain sub-groups of users.

The complement to this top-down approach is a *bottom-up, use case* approach. EuroFOT has defined a use case as “a specific event in which a system is expected to behave according to a specified function.” An example would be car following.

Events are affected by *situations*, where a situation is “one specific level or a combination of specific levels of situational variables.” An example would be the combination of rainy weather, darkness and motorway driving. Finally a *scenario* is “a use case in a specific situation”, and thus might be car following in the rain and the dark on a motorway.

It can readily be seen that the use case approach provides a mechanism to drill down into quite detailed and specific scenarios about system operation and user interaction with a system. This complements the much more broad-brush viewpoint afforded by the six areas. However, there is a danger that the number of scenarios can explode with consequent demands on data collection procedures, and therefore a process of prioritising hypotheses is recommended.

Road-testing the procedure

The process originally recommended by FESTA has subsequently been applied in a European networking project called FOT-Net. This project is intended to provide a platform for knowledge exchange and for strategic networking among existing and future national, European and international Field Operational Tests.

In late 2008, FOT-Net held a seminar on the initial stages of carrying out a Field Operational Test. In small discussion groups, hypotheses were formulated about the potential impacts of three imaginary systems, all of which were conceivable with current technology. Each group used both the top-down and the bottom-up approach.

It was concluded from the seminar that the bottom-up and the top-down approach are complementary. On its own, neither approach produces the full set of hypotheses generated by the combined approach. It was also concluded that a group setting is necessary for generating useful hypotheses and for prioritizing them. It is not work that can be done by a single person or by remote communication, such as via email.

A subsequent FOT-Net seminar on combinations of systems and functions was held early in 2009. There it was agreed that defining hypotheses for combined functions is difficult, complex and time-consuming. The seminar recommended a two-step approach in which first hypotheses are identified for each function in isolation and then the discussion moves on to the effect of combining functions.

In dealing with combinatory effects, a useful approach is to consider whether the effects are additive or whether there is an interaction in the statistical sense. There may be unexpected interactions between systems despite entirely different functionalities and purposes. Such interactions can be both in terms of technical operation of systems and in terms of user impacts. Finally, it is clear that hypotheses about combined effects should influence the experimental design.

Further outcomes of this seminar was that there was a need to find an appropriate balance between costs and budget, and science and practicality. Hence prioritizing of research questions and hypotheses was vital. A detailed analysis of hypotheses would lead to appropriate events to study and to definitions of performance indicators. The development of experimental design was itself iterative starting with a first rough design and moving on to further refinements.

Conclusions

It is not possible to claim that arriving at a set of hypotheses is an entirely objective exercise or that a different team would arrive at the same set of hypotheses. There is always an element of subjectivity and intuition in hypothesis creation. But a structured approach is much more likely to ensure coverage of a wide range of potential impacts than an unstructured one. Iteration between a global top-down approach and a bottom-up use case approach has proven its value, by generating hypotheses that neither approach produced alone. Finally, insistence on stating a specific mechanism linking cause and effect ensure that hypotheses are grounded in theory and are not just wild guesses.

References

- Draskóczy, M., Carsten, O., & Kulmala, R. (1998). *Road safety guidelines*. Deliverable B5.2 of CODE project (TR1103). Birmingham, UK: Atkins Wootton Jeffreys.
- Elvik, R. (2008). Criteria for causal inference: lessons from epidemiology applied to safety evaluation. *Workshop on future directions in highway crash data modelling*. Washington D.C.: Transportation Research Board, National Academy of Sciences.
- FESTA Project (2008). FESTA Handbook Version 2. Deliverable D6.4 of Field operational test support Action (FESTA). Available at <http://www.its.leeds.ac.uk/festa/>.
- Markus, K.A.(2008). Hypothesis formulation, model interpretation, and model equivalence: implications of a mereological causal interpretation of structural equation models. *Multivariate Behavioral Research*, 43, 177-209.
- Michon, J.A. (1985). A critical review of driver behaviour models. In L. Evans and R.G. Schwing (Eds.), *Human behavior and traffic safety* (pp. 485-520). New York: Plenum Press.
- VTTI (2007). *Development and prioritization of research questions: preliminary results*. Task S05 of SHRP2 Research Programme. Blacksburg, Virginia, USA: Virginia Tech Transportation Institute.