

6th Transport Research Arena April 18-21, 2016



Methodology for field operational tests of automated vehicles

Yvonne Barnard ^{a,b,*}, Satu Innamaa ^c, Sami Koskinen ^c, Helena Gellerman ^d,
Erik Svanberg ^d, Haibo Chen ^a

^a*Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK*

^b*ERTICO – ITS Europe, Avenue Louise 326, Brussels, B-1050, Belgium*

^c*VTT, Espoo / Tampere, P.O. Box 1000, 02044 VTT, Finland*

^d*SAFER, Lindholmospiren 3, Göteborg, SE-402 78, Sweden*

Abstract

Over the past decade a large number of Field Operational Tests (FOT) have been conducted to test Intelligent Transport Systems (ITS) in real traffic conditions with thousands of drivers. In order to ensure scientifically sound studies a FOT methodology was developed in the FESTA project. Currently we are on the brink of a new series of large scale FOTs, testing automated and autonomous vehicles. A common FOT methodology serves the following purposes: (1) to ensure that a systematic and scientific approach is taken by FOTs, (2) to enable the assessment of the impact of large-scale introduction of ITS on safety, mobility, efficiency and environment, (3) to be able to compare results of different FOTs, and (4) to build a community and facilitate knowledge exchange. FESTA focuses strongly on the drivers of vehicles, and the changes in their behaviour when driving a vehicle that is instrumented with new systems. In FESTA, it is recommended that driving with an ITS is compared with driving without it (the baseline). However, what will be the focus of the new FOTs? And what will be the main research questions these FOTs will address? And what is the baseline? Three types of focus can be distinguished; centred on the user, the vehicle or the context. In this paper we discuss the requirements for a methodology that addresses these three types of focus. We investigate how the current FOT methodology may be adapted or may need to be completely changed. Special attention is given to the type of data that is needed for baselines and for answering research and impact questions.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of Road and Bridge Research Institute (IBDiM)

* Corresponding author:

E-mail address: y.barnard@leeds.ac.uk

Keywords: Automated driving; Field Operational Tests; evaluation; methodology; road transport

1. Introduction

Nomenclature

| | |
|-----|---------------------------------|
| FOT | Field Operational Test |
| ITS | Intelligent Transport System |
| SAE | Society of Automotive Engineers |

1.1. Background

Over the past decade a large number of Field Operational Tests (FOT) has been conducted in Europe, the US, Japan, Australia, and other countries to test Intelligent Transport Systems. The European Commission has sponsored several large-scale FOTs (Barnard et al., 2015). Advanced driver assistance systems including cooperative systems, (with communication between vehicles or between vehicles and infrastructure) have been tested with thousands of drivers in real traffic conditions. Examples of these FOTs were euroFOT (Kessler et al., 2012), TeleFOT (Mononen et al., 2013), DRIVE C2X (Schulze et al., 2014) and FOTsis (Alfonso et al., 2015).

In 2007, the European Commission launched a call for proposals to develop a FOT methodology before large-scale European FOTs would be funded. The FESTA consortium that was granted this project consisted of a large number of stakeholders, from industry, research, public authorities, and user and network organisations. FESTA produced guidance and support for the upcoming FOTs. The FOTs had as an objective to comprise a comprehensive program of research to assess the impacts of ICT systems on driver behaviour, both in terms of benefits for drivers (e.g. more comfort and increased safety) and of larger scale socio-economic benefits (e.g. less congestion and fewer accidents) (Barnard & Carsten, 2010). A handbook was developed with many practical recommendations (FESTA, 2014; Regan & Richardson, 2009). The basis of this handbook was a methodology, to be followed by the FOTs in order to ensure scientifically sound studies and allowing comparability between FOTs (Carsten & Barnard, 2010). Since 2008 this methodology has not only been adopted by FOTs funded by the European Commission but also by many nationally (or otherwise) funded projects, and has influenced FOTs outside Europe. The methodology has been regularly updated by the FOT-Net support actions, taking into account the lessons-learned (www.fot-net.eu).

The FESTA methodology is summarised in Figure 1. There are several steps, which although described in a linear way, are performed in iteration. The V-shape (fig.1) shows the dependencies between the different steps in the left- and right-hand side of the V. The steps can be summarised as:

- Defining the study: Defining functions, use cases, research questions and hypotheses
- Preparing the study: Determining performance indicators, study design, measures and sensors, and recruiting participants
- Conducting the study: Collecting data
- Analysing the data: Storing and processing the data, analysing the data, testing hypotheses, answering research questions
- Determining the impact: Impact assessment and deployment scenarios, socio-economic cost benefits analysis

A common FOT methodology had to ensure that a systematic and scientific approach was adopted by the FOTs. Performing large-scale FOTs was not only meant to enable the assessment of a single or a few systems but to get a better view of the potential impact of large-scale introduction of ITS on safety, mobility, traffic efficiency and the environment. These societal challenges are complex, and require multiple studies that all contribute to a better understanding and assessment of how transport may become more intelligent, and how it may influence society. Providing a common general approach and a common vocabulary makes it easier to compare studies on similar systems, to gain a better understanding of the changes in road-users' and traffic behaviour, and to interpret outcomes.

To perform FOTs, a wide range of expertise is needed, such as on technical, legal and ethical, human factors, statistical, computer-science, political and organizational aspects. FESTA was developed and subsequently maintained and updated by a large number of experts from some 30 organizations. The methodology and handbook is more than a theoretical and practical text, it is the focal point of an international community involved in performing the studies and using their outcomes. The three FOT-Net support actions funded by the European Commission since 2008 (FOT-Net, FOT-Net 2, and FOT-Net Data: www.fot-net.eu) ensure the building and sustaining of this community, by workshops, meetings, seminars, documentation, a website, and a wiki in which a large number of FOTs are described (wiki.fot-net.eu; Barnard et al. 2014). This supports the exchange of experiences and lessons-learned, making the methodology a living one.

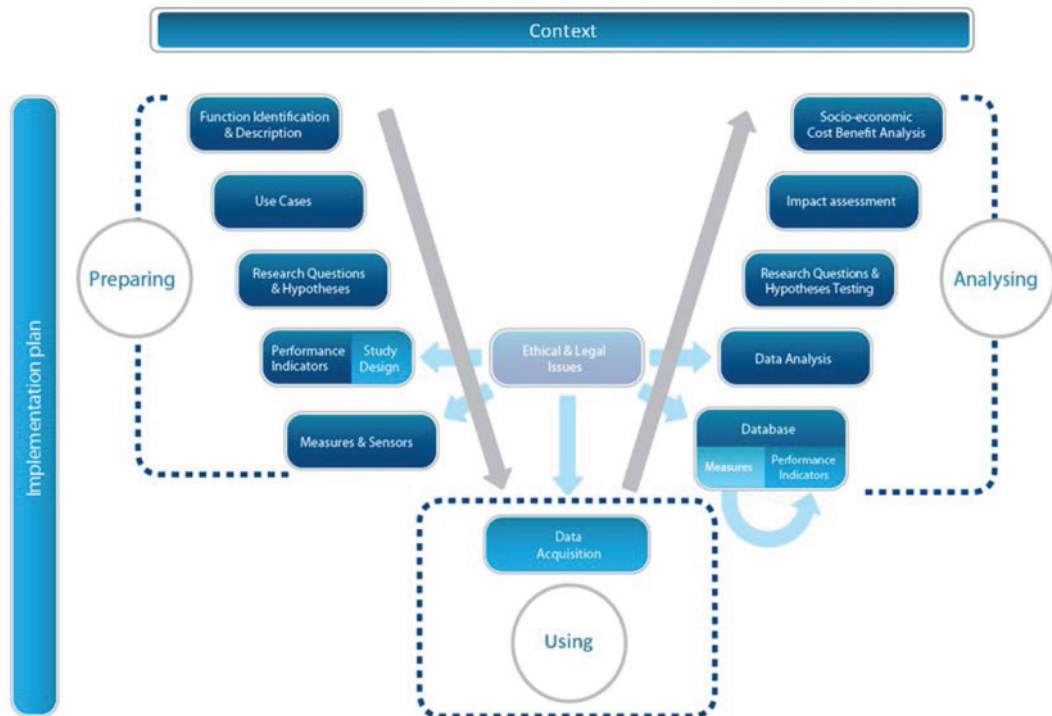


Fig. 1. The FESTA methodology.

1.2. Purpose of this paper

In the coming years we will see a new series of large FOTs, on (partly) automated vehicles. The term automated vehicles may refer to several levels of automation, ranging from vehicles with systems that may take over some functions from the driver to completely driverless vehicles. The SAE distinguishes five levels of automation, from driver assistance to full automation (SAE, 2014). If we talk about the highest levels of automation, it may be better to use the term “user” instead of “driver”, as there is no longer a distinction between active drivers and passive passengers. This paper discusses how FOTs on automated vehicles will be conducted, whether the FESTA methodology will still be of use, whether we need to adapt it, and whether we need a new methodology and a new research paradigm.

FESTA is strongly centred on the drivers of vehicles, and the changes in their behaviour when driving a vehicle that is instrumented with new systems. In FESTA it is recommended that driving with the system is compared with driving without it (the baseline). However, what will be the focus of the new FOTs? And what will be the main questions these FOTs will address? What is the baseline? And how do we determine the impact?

Before we discuss these questions two things should be noted. Firstly, FOTs are not the only way to evaluate automation, it is a method that is complementary to (or is preceded by) simulator, laboratory and/or simulation studies, expert assessments, technical tests on test-tracks, public acceptance studies, etc. Secondly, FESTA is not concerned with the technical testing of systems; it assumes that the systems are stable prototypes or ready-to-market. When large-scale FOTs will be conducted with automated vehicles we may also assume that the vehicles concerned have been tested technologically and are ready to drive on real roads.

2. Goals and focusses of automation FOTs

Three types of goals and focusses can be distinguished for automation FOTs:

1. **User centred tests** addressing questions about how the user/driver reacts to automation and uses it, how (s)he understands the capabilities of the vehicle, what (s)he will do when not supervising the surrounding vehicles/traffic, the transition from automated driving back to the driver during the journey, drivers' situation awareness, the interaction between automated vehicles, their drivers and other road users including vulnerable road users and drivers of conventional vehicles, user acceptance, changes in mobility patterns, etc. These questions are specifically of interest for vehicles that are not fully automated (especially SAE 2-3(-4)), and where there remains a role for the driver. The results could be compared with FOTs in which a lesser (or even no) level of automation was used (SAE 0-1).
2. **Vehicle centred tests** addressing the question of how the automated vehicle behaves in different traffic conditions. Here questions about the situational picture that the automated vehicle creates, and the interaction of the automated vehicle with other automated and conventional vehicles, other road users and with the infrastructure need to be answered. The baseline could be the (spectrum of) behaviour of conventional vehicles. However, this criterion is not straightforward as driving cultures differ from one location to another and over time. The tests should cover a variety of driving conditions, levels of automation and penetration rates. The traffic flow of mixed traffic of lower levels of automation and of mainly fully automated vehicles may have different dynamics. There may appear new types of traffic accidents and conflicts. What will happen if traffic management or automatic speed control determine the behaviour of vehicles?
3. **Context centred tests** addressing questions of how mobility changes, how this affects mobility services, what the impacts are on traffic flow level or on transport system level, what the position is of other road-users like pedestrians and cyclists, what ethical choices might be involved, and what would be the impacts on the built-up environment and society. These types of question are extremely important but not easy to investigate with FOTs as these impacts typically take a longer period of time to evolve than the duration of a FOT and, certainly, these changes do not take place with the penetrations that a FOT is able to put up.

Examples of types of research questions and topics for FOTs with different focus are summarised in Table 1.

Table 1. Examples of research topics for different focusses.

| Focus on: | Example research topics on: |
|---------------------|---|
| User | Drivers' reactions to automation, understanding of its capabilities Ways of using automation Drivers' activities when not supervising the vehicle or surrounding traffic Transition from automated driving back to driver during the journey Situation awareness and its recovery Interaction between automated vehicles, their drivers and other road users Changes in personal mobility patterns User acceptance |
| Vehicle | Vehicle behaviour in different traffic situations Situational picture that the automated vehicle creates Interaction with other vehicles, road users, infrastructure Traffic flow dynamics Accidents and conflicts |
| Context and society | Changes in mobility and mobility services Transport system level changes Position of other road-users like pedestrians and cyclists Ethical choices Impact on the build-up environment and society |

3. The baseline

In FESTA it is recommended that driving with the system is compared with driving without it (the baseline). For fully automated vehicles we no longer have a system that can be viewed as independent from the vehicle itself, the whole vehicle is now the system. Still, we want to evaluate something that is done typically in relation to something else. Below we will discuss several options for a baseline.

It might be ideal if we could compare automated driving with non-automated driving (SAE 0). However, there are several reasons why this is not so easy. In the first place, some current vehicles already have some automated functionalities and can be considered to belong to SAE levels 1 or 2. Thus, a traffic flow of 100% non-automated (SAE 0) vehicles is even now only theoretical or belongs to the past. As vehicles and traffic evolve also in other ways than just in relation to automation, comparison to something from the past does not give the true impact compared to the current or future situation.

If the baseline is chosen to be the prevailing vehicle population of the time when the automation FOT is conducted, we will always be comparing different kinds of baseline and a meta-analysis will be difficult to perform. If this option is selected, an interesting approach could be not to establish a baseline during the FOT, but to look at data and results from Naturalistic Driving Studies. In these studies "normal" everyday driving is studied (Eenink et al. 2014). This behaviour could be compared with the participants' behaviour driving in automated vehicles.

If a theoretical traffic flow for a certain level of automation is selected as baseline (any automation level lower than the tested vehicle may be selected), attention must be paid to selecting the automated functionalities of the baseline vehicle or vehicle flow that are meaningful for the impact assessment purposes. A vehicle belonging to certain level of automation may be very different from another vehicle belonging to the same automation level if different aspects of driving have been automated. Thus, if parking behaviour of automated vehicles is studied, the automated parking functionalities are the key criterion in selecting the baseline.

If the comparison is made on a one-to-one level, the baseline vehicle may be a conventional vehicle or a vehicle with lower level automation. However, when impacts are assessed on traffic flow level, the penetration of different levels of automation is a key issue, as the penetration of connected or autonomous vehicles as well as the connectivity may impact on, for example, the car-following behaviour of vehicles. One could compare situations where there is a high number of automated vehicles in a certain area, versus a low number or just one. This may require a more experimental set-up, bringing vehicles together, like a study on platooning would do.

When selecting research questions and hypotheses and building a test design, we must bear in mind that things should not be compared when they are not comparable. For example, the percentage or duration of eyes-off-the-road may be much higher with automation than in conventional vehicles, but is this of any interest? Thus, the baseline should be always selected wisely and guidance from a methodology would be useful

Maybe the whole idea of a baseline is not always suited for automation FOTs. Some forms of automation, like autonomous or driverless vehicles, mean a radical change in transport. Comparison with the “old” situation may not be very useful, and studying new emerging patterns may be of more interest. Studying effects against a baseline is important if decisions have to be made on whether the introduction of new systems is desirable. If automation is seen as a process that will continue anyhow, a baseline becomes less important. Similar questions occurred with the introduction of nomadic devices. It is not so interesting to compare the behaviour of travellers without and with a smartphone as their penetration rate went up so quickly that comparison became meaningless, other than for historic reasons.

4. Studying impact of automation

In Figure 2, the FESTA approach to impact analysis is illustrated showing how the impacts on safety, mobility and environment can be assessed as well as cost-benefit ratios. Note that the impact analysis is focused on areas about which FOTs can provide valuable insights. When we talk about automation, other societal impacts will also be relevant, such as land-use, urban planning, employment, transport poverty, etc. These are not elements a FOT can measure. Still, it would be good to try to look at wider and new impact areas, and to start a discussion on whether the current methods and indicators are broad enough to answer questions about societal challenges.

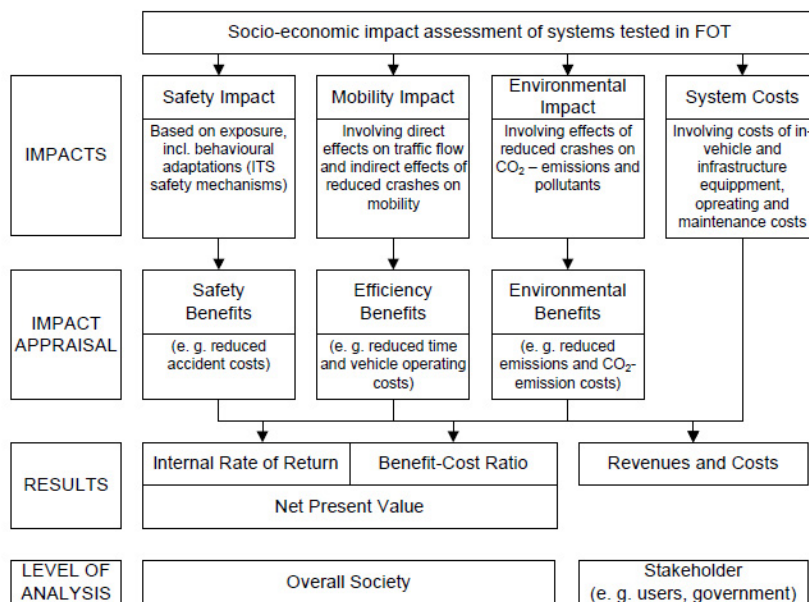


Fig. 2. Impact analysis according to FESTA, taken from FESTA, 2014.

One of the main challenges for impact analysis will be the scaling up. Where it is relatively easy to perform a scaling up exercise for one Advanced Driver Assistance System (“How many unintended lane-departures could be avoided if X percent of the cars had a Lane Departure Warning system?” or “How many lives could be saved annually on European level?”), scaling up with statistical methods for automated cars is not straightforward. Scaling up in more complex situations is usually done by using macroscopic traffic models or mathematical models. These models will have to be revised to cater for automation. One of the problems is that the whole traffic situation and the transport and mobility behaviour of people might change drastically if the use of automated vehicles will become widespread.

For automation, different scenarios are envisaged (Sessa et al., 2013). Some foresee a gradual change, with more and more vehicles being equipped with higher levels of automation over the coming years. Others focus on scenarios where (semi)automated vehicles will be on the road as a new type of vehicle, bringing a revolution to road transport. Performing a socio-economic impact analysis of automated vehicles will have to take into account these different scenarios. New mobility patterns and services will have an impact on these but the time-frame of these impacts is longer than the duration of a FOT, making them difficult to measure.

In order to be able to perform a cost-benefits analysis it is necessary to monetize all the foreseen effects. For the traditional costs and benefits (costs of accidents, time lost due to congestion, etc.) this may still be possible, however, for new performance indicators and impact areas this may require new assumptions and methods to perform the calculations. In addition, traditional measured costs like the cost of time spent in congestion may have to be revised if the time is no longer just wasted waiting to get to the destination but can be utilised e.g. by processing emails or in a conference call while the (fully) automated vehicle takes care of driving.

As evaluation of automated vehicles is no longer at the level of separate systems, it will be necessary to look at the interaction between impact areas, and the integration of results, in terms of costs and benefits, but probably also in terms of societal changes.

Internationally there is a large interest in impact assessment, a tri-lateral working group between US, EU and Japan is working on the harmonization of impact assessment framework for automation in road traffic. The U.S. Department of Transportation has recently published a report presenting a framework for estimating the potential benefits and dis-benefits of technologies contributing to the automation of the US’ surface transportation system (Smith et al. 2015).

5. A new methodology for automation FOTs?

The question we set out at the beginning of this paper is whether a new methodology for performing automation FOTs is necessary or whether we can adapt the existing one. Before we try to answer this question we should ask whether a common methodology is necessary at all. The reasons for having one for conventional FOTs were to ensure a systematic and scientific approach, to enable the assessment of the impact of large-scale introductions, to be able to compare results and perform meta-analysis, and to build a community of FOT professionals and to facilitate knowledge exchange and mutual learning among them. These reasons remain valid for automation FOTs, and maybe even more so.

To be able to profit from automation FOTs, sharing of data is important. Sharing data is not evident, because of competition, privacy issues, and the costs of making data available to others. However, if data that are not sensitive (from a commercial and personal point of view) are shared and sharing is facilitated, we will be able to learn much more, and much more efficiently, about the impacts automation may have. The FOT-Net Data support action has developed a data sharing framework to facilitate this sharing (Barnard et al, 2014, Gellerman et al., 2015).

All over Europe, and internationally, initiatives are taken to bring automated vehicles onto real roads to test them. Competition between manufacturers may be fierce. Automation may bring lots of benefits for transport, but also many challenges are foreseen, and there is no solution for every problem yet. For example, the question of the role of the driver (or the question of whether there will still be a responsible driver) or who is responsible if (/when) something goes wrong leads to heated discussions. Can we leave everything to car-industry to solve, and/or should authorities support the introduction of automation in road traffic? If so, how, where and based on what? We are entering non-charted territory, and we will need to learn as much as possible from each other’s experiences, gaining as much knowledge as possible about what automation would mean for society. Policy makers need sound evidence of the consequences before they are able to make decisions on regulation for automation, and transparency in the process of how the conclusions from studies were reached. Where competition is needed for the development of innovative

systems and vehicles, collaboration is needed for ensuring that automation does not have detrimental effects, and that the impacts are established in an honest and sound way. Finding the right balance between openness and (commercial) confidentiality was always an issue in the large-scale European FOTs, and will be a big challenge for future FOTs.

Now, if we agree on common methodology, the question remains whether we should just adapt the old FOT methodology or devise a new one. If we look at the current FESTA methodology, many steps will still be valuable for automation FOTs. In the first place the scientific approach, driven by research questions and hypotheses, will still be the best way to gain evidence. In the second place many of the more practical issues of setting-up and conducting these studies will remain the same (at least on a high level). The experiences gained with instrumentation, practical issues in organising FOTs, and data analysis, as put down in the FESTA handbook, will remain valid, and of great use to new studies. Some practical issues will become more urgent, for example, the collection and use of a huge set of real-time data that will be generated and used by automated vehicles. However, the existing sensors and other equipment in automated vehicles may make instrumentation of test vehicles for FOT purposes easier or more cost-efficient than before.

Nevertheless, new issues need to be addressed. As can be seen in the FESTA V in Figure 1, establishing the context is the step that steers the whole process. Automation will give rise to new contexts, such as the scenarios for the introduction of automation as discussed earlier. Where questions about legal and ethical issues were already difficult to answer in FESTA, automation will bring a whole new and difficult set of issues, such as the responsibility of drivers, users and owners of automated vehicles. New methods to assess the longer-term impacts than those visible during the FOT life-cycle need to be developed and, preferably, harmonised.

FESTA has a strong focus on driver centred studies (see Section 4), and the other focusses (vehicle and context) need to be addressed as well, probably requiring new concepts and definitions. We already discussed the change in relevant impact areas, and need methods to perform impact assessment.

We did not provide the answer in this paper as to whether a new methodology for performing automation FOTs is necessary or whether we can adapt the existing one, and we raised more questions than we answered. It is not up to the authors of this paper to provide all the answers. FESTA is a living methodology, developed, adapted and adopted by a large community, based on consensus and sharing experiences. This will also be the case with a methodology for automation FOTs. However, we hope to have made a convincing case for a common methodology, keeping the valuable approach and elements of FESTA, and developing new methods to answer new questions, and to bring knowledge on new focus points of the automation FOTs.

Acknowledgements

FOT-Net Data is a Support Action in the seventh Framework Programme Information and Communication Technologies. It is funded by the European Commission (EC), DG Connect, Grant Agreement number 610453. In addition, we acknowledge the discussions in 'US-EU-JPN Harmonization of Impact Assessment Framework for Automation in Road Traffic' Working Group.

References

- Alfonso, J., Sánchez, N., Menendez, J.M., Cacheiro, E., 2015. Cooperative ITS communications architecture: the FOTsis project approach and beyond. *IET Intelligent Transport Systems*, 9 (6); pp. 1-8.
- Barnard, Y. (Ed.) 2014a. FOT-Net 2 Deliverable D2.2 Report on FOT Network activities.
- Barnard, Y., Carsten, O., 2010. Field Operational Tests: challenges and methods. In: Kreams, J., Petzoldt, T., Henning, M. (Eds.): *Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems*, pp. 323-332. Lyon: HUMANIST publications.
- Barnard, Y., Fischer, F., Flament, M., 2015. Field Operational Tests and Deployment Plans. In: Campolo, C., Molinaro, A., Scopigno, R. (Eds.). (2015). *Vehicular ad hoc Networks: Standards, Solutions, and Research*. Springer, pp. 393-408.
- Barnard, Y., Koskinen, S., and Gellerman, H., 2014. A platform for sharing data From Field Operational Tests. In: *Proceedings of the 21st ITS World Congress 2014 in Detroit*.
- Carsten, O., Barnard, Y., 2010. Preparing field operational tests for driver support systems: a research-oriented approach. In de Waard, D., Axelsson, A., Berglund, M., Peters, B., Weikert, C. (Eds.) *Human Factors: A system view of human, technology and organisation*. Maastricht, the Netherlands: Shaker Publishing. pp. 243 – 248.

- Eenink, R., Barnard, Y., Baumann, M., Augros, X., Utesch, F., 2014. UDRIVE: the European naturalistic driving study. Proceedings of the Transport Research Arena, Paris.
- FESTA, 2014. FESTA Handbook, Version 5. Available at: http://wiki.fot.net.eu/index.php?title=FESTA_handbook.
- Gellerman, H., Svanberg, E., Bärghman, J., Barnard, Y., 2015. Data Sharing Framework for Naturalistic Driving Data. Proceedings of the 22nd World Congress on Intelligent Transport Systems in Bordeaux.
- Kessler, C., Etemad, A., Alessandretti, G., Heinig, K., Selpi, Brouwer, R., Cserpinszky, A., Hagleitner, W., Benmimoun, M. 2012. Final Report. euroFOT Deliverable D11.3. Available at: http://www.eurofot-ip.eu/en/library/deliverables/sp1_d113_final_report.htm
- Mononen, P., Franzen, S., Pagle, K., Morris, A., Innamaa, S., Karlsson, M., Touliou, K., Montanari, R., Fruttaldo, S., 2013. Final Report, TeleFOT Deliverable D1.15.
- Regan, M. A., Richardson, J. H., 2009. Planning and implementing field operational tests of intelligent transport systems: a checklist derived from the EC FESTA project. *Intelligent Transport Systems, IET*, 3(2), pp. 168-184.
- SAE 2014. Automated driving. SAE international standard J3016. Available at: http://www.sae.org/misc/pdfs/automated_driving.pdf
- Schulze, M., Mäkinen, T., Kessel, T., Metzner, S., Stoyanov, H., 2014. Final Report, DRIVE C2X Deliverable D11.6. Available at: <http://www.drive-c2x.eu/publications>
- Sessa, C., Fioretto, M. et al., 2013. Blueprint of alternative city cyber-mobility take-up scenarios. CityMobile2 Deliverable D27.1.
- Smith, S., Bellone, J., Bransfield, S., Ingles, A., Noel, G., Reed, E., Yanagisawa, M., 2015. Benefits Estimation Framework for Automated Vehicle Operations. Report FHWA-JPO-16-229. John A. Volpe National Transportation Systems Center, U.S. Department of Transportation. Available at: http://ntl.bts.gov/lib/55000/55400/55443/AVBenefitFrameworkFinalReport082615_Cover1.pdf