

Can We Trust Driver Behaviour Assessment?

Examples from Research in Simulators and in the Field

Katja Kircher, Christer Ahlström

The Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden

Introduction

It is very common to compare mean values of driving performance indicators (PI) like mean speed, the standard deviation of lateral position, time headway, mean glance duration, and many more, in order to investigate possible differences between different treatment groups. Just like all the PIs mentioned here, most of them describe aspects of the control level of driving behaviour according to Michon's control hierarchy [1]. When means differ significantly between treatment groups, this is often interpreted in relation to traffic safety gains or losses.

In this paper we are going to discuss possible pitfalls with the use and interpretation of such PIs, based on examples from different studies. Finally, we suggest a number of possible solutions to avoid some of the issues discussed here.

Individual differences

As people may perform differently under similar conditions, it is common to compare the mean values of groups to find out whether the groups differ systematically in their performance. How the individuals differ within the groups, and what the possible reasons for those differences are, is often neglected. This is illustrated with an example from a study conducted in two different driving simulators (Figure 1). In addition to a baseline drive the participants had to drive on a relatively curvy rural road while either performing a mostly visual, a solely cognitive, or a mostly haptic task. They did this once under car-following and once under free driving conditions. The additional tasks were presented on three different levels of difficulty. The participants were instructed to complete those tasks as best they could, while still having in mind that they were driving, that is, the main focus was directed at the additional task [2].

Several aspects can be noticed by taking a closer look at **Error! Reference source not found.**, which shows the time in seconds a driver spent with at least two wheels in the oncoming lane. The total time driven per condition was 30 s. The thin lines stem from individual drivers. Especially for the visual condition it is noticeable how much the values vary between participants. What is more interesting, they do not only vary in a quantitative manner, with one participant getting into the oncoming traffic more than the other, but also in a qualitative manner. Except for the visual condition, most participants do not leave their lane at all, or for less than 2 s, as can be deducted from the median values. Some other participants, however, drive with at least two wheels in the oncoming lane for a third of the time or more. This may be interpreted as some participants losing control, or consciously accepting a drift into the oncoming lane, while others manage to keep the vehicle within the lane boundaries. With those data and interpretations as background, it does not seem meaningful to analyse mean values for PIs like this one. Instead it might be much more educative to investigate the underlying factors that might lead to the quantitatively different individual results.

Systematic differences

Especially when comparing the results from different studies, systematic differences between those studies can influence the results in unforeseen ways. These differences can for example be coupled to participant selection, the test bed in use, or the way the PI was computed.

It has already been indicated by Greenberg et al. [3] that motion cueing may have a strong impact on lateral driving PIs when disturbances, like the execution of additional tasks, are present. A number of further studies

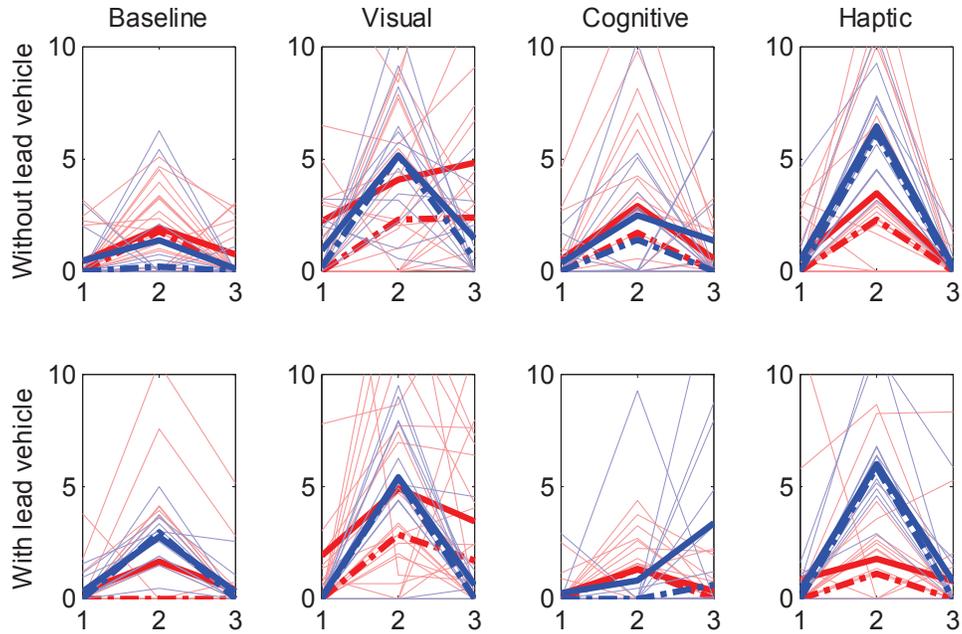


Figure 1. Individual, mean (solid lines) and median (dashed lines) lane departure time results (in s) for driving without (upper row) or with (lower row) a lead vehicle under baseline conditions, or with a visual, cognitive or haptic additional task (from left to right) in a high fidelity moving base simulator (red) or a static simulator (blue).

corroborate the fact that the type of motion cueing influences results [4,5,6]. Therefore, when comparing results between different simulators it has to be kept in mind that the measured Pis are not necessarily directly comparable.

Participant recruitment is another factor that may influence results. It is not uncommon that universities recruit participants amongst students, as this is convenient. In the study mentioned above, the participants who drove in the moving base simulator were recruited from the general public, whereas the participants for the fixed base simulator were recruited from engineers who were employed at the company running the study. The participants had also been asked to perform the additional tasks while standing still, that is, in this condition the additional task was the only task performed, so simulator differences were eliminated. For all three modalities tested, the engineers clearly and significantly outperformed the members of the general public. This difference in performance between groups in the additional task may have had a systematic impact on the capability to perform the driving task, such that studies with different types of participants may lead to qualitatively or quantitatively differing results.

Furthermore, a PI used in different studies under the same name can be computed in different ways. To illustrate this, we focus on the window size that is used to compute the PI “standard deviation of steering wheel angle”, an indication of how variably the driver turns the steering wheel over a certain period of time. The data used in the example stem from a field study, in which the drivers used an experimental car as their own during the period of a month [7]. With the help of the distraction detection algorithm *AttenD*, which uses eye tracking data as input, the drivers were classified as either distracted or attentive for certain time windows [8,9]. Depending on the time window used to compute the “standard deviation of steering wheel angle” there was either no significant difference or there was a significant difference between distracted and attentive drivers (**Error! Reference source not found.**). Relating back to what was discussed above, the graph containing the standard deviation of the PI in the right part of **Error! Reference source not found.** illustrates once more how wide the behavioural range is within each group, in spite of the significant differences in mean values for some of the window sizes.

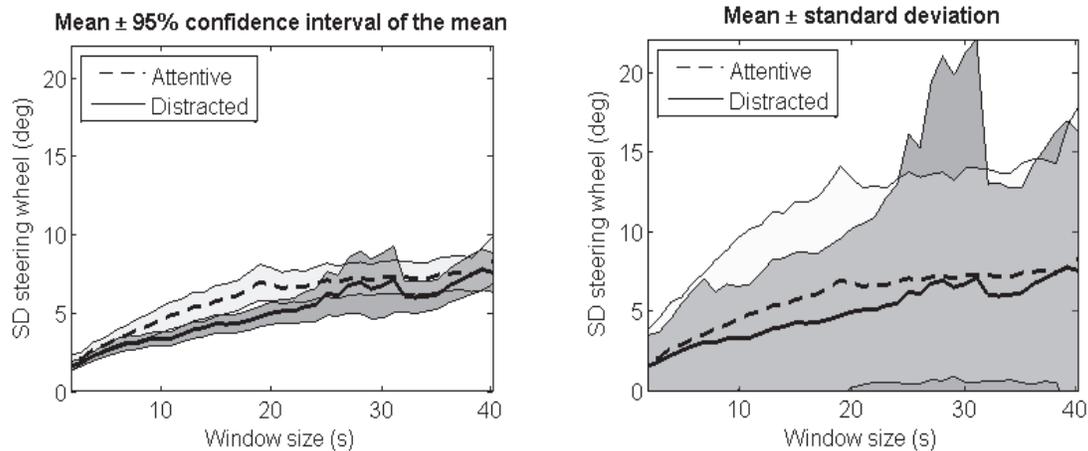


Figure 2. The PI “standard deviation of steering wheel angle” computed for data logged in the field for window sizes between 1 s and 40 s. In the graph on the left the mean with its 95 % confidence interval and in the graph on the left the mean with one standard deviation is shown.

What do performance indicators really indicate?

When it is found that on average the standard deviation of lateral position (SDLP), that is, how much a vehicle sways around its medium lateral position in the lane, increased with, say, 12 cm, what does this tell us? Obviously, the trajectory of the vehicle is somewhat less straight. Can this be linked directly to a safety hazard, though? We argue that more information is necessary before such an interpretation can be made. We would need to know, for example, whether the vehicle actually leaves the lane, whether other traffic is present, whether the other traffic comes dangerously close due to the increased swaying, whether the driver is aware of the increased swaying, whether he accepts it in the situation at hand, and so on.

Any difference in means can become significant when enough participants are used, such that the size of the measured effect plays an important role. In addition to the purely statistical effect size, which tells us how much the mean values differ in relation to the standard deviation within the population, the “practical effect size” should be considered as well. As mentioned in the example above, which practical implications does an increased SDLP have for safety, road construction, etc.? On a wide and straight road, with an average lateral position in the middle of the lane, the 12 cm increase may not have any safety implications at all. On a narrow road, however, this could mean the difference between crossing over into the oncoming lane or not. Then again, a driver may use the oncoming lane on purpose, when cutting the curves on a rural road. Obviously, the driver would not do so when oncoming traffic was present, such that an increased SDLP as a consequence of intentional behaviour does not need to be correlated to changes in the level of traffic safety at all.

Possible solutions?

It is very difficult to link this kind of PIs to actual crashes. A large-scale attempt will be made during the SHRP II programme, with 2000 instrumented vehicles and an even higher number of drivers. However, even studies of that scale will not be able to answer all the questions of what changes in driving behaviour related PIs really mean. Therefore we suggest a number of pragmatic measures to enhance the safety related validity of PIs used in studies.

Instead of mainly using PIs that describe the control level of driving behaviour, it is suggested to develop indicators that describe the tactical level. These are expected to be easier to understand and interpret. It is also assumed that the tactical level is affected by external factors sooner than the control level is, which would make

such PIs more sensitive. A drawback is the fact that tactical PIs may have to be adapted to the present situation much more, which makes them harder to standardise for comparisons between studies.

PIs should have a link to violations or other behaviour that already has been linked to an increased risk in traffic. Of course, PIs can also be used to estimate efficiency or other aspects of the traffic system, but then a link to those concepts should be documented.

In order to incorporate the qualitative/quantitative aspect of behavioural differences it may be meaningful not only to evaluate the difference in mean values between groups, but to focus on the percentage of drivers who are able to perform a task within given boundaries, for example, to send an sms within a certain amount of time while remaining within one's lane and within a certain speed interval.

References

1. Michon, J. A. (1985). A Critical View of Driver Behavior Models. What Do we Know, What Should we Do? In L. Evans & R. Schwing (Eds.), *Human Behavior and Traffic Safety*. New York: Plenum Press.
2. Kircher, K., Ahlstrom, C., Rydström, A., Ljung Aust, M., Ricknäs, D., Almgren, S., et al. (2012, in press). *Secondary task workload test bench - 2TB. Final report*. Linköping, Sweden: The Swedish National Road and Transport Research Institute.
3. Greenberg, J., Artz, B., & Cathey, L. (2003). *The effect of lateral motion cues during simulated driving*. Paper presented at the DSC North America.
4. Correia Grácio, B. J., Wentik, M., Valente Pais, A. R., van Paassen, M. M., & Mulder, M. (2011). Driver behavior comparisons between static and dynamic simulation for advanced driving maneuvers. *Presence: Teleoperators and Virtual Environments* **20**(2), 143-161.
5. Feenstra, P., van der Horst, R., Correia Grácio, B. J., & Wentik, M. (2010). Effect of simulator motion cueing on steering control performance. Driving simulator study. *Transportation Research Record: Journal of the Transportation Research Board* **2185**, 48-54.
6. Valente Pais, A. R., Wentik, M., Van Paassen, M. M., & Mulder, M. (2009). Comparison of three motion cueing algorithms for curve driving in an urban environment. *Presence: Teleoperators and Virtual Environments* **18**(3), 200-221.
7. Kircher, K., Kircher, A., & Claezon, F. (2009). *Distraction and drowsiness. A field study* (VTI Report No. 638A). Linköping: VTI.
8. Kircher, K., & Ahlstrom, C. (2009). Issues related to the driver distraction detection algorithm AttenD. *First International Conference on Driver Distraction and Inattention*.
9. Kircher, K., & Ahlstrom, C. (2010). Predicting visual distraction using driving performance data. *Association for the Advancement of Automotive Medicine Annual Scientific Conference*, Las Vegas, NV, USA.