ON THE ASSESSMENT OF CRITERIA FOR DRIVER IMPAIRMENT; IN SEARCH OF THE GOLDEN YARDSTICK FOR DRIVING PERFORMANCE

Karel Brookhuis and Dick de Waard
University of Groningen
Groningen, the Netherlands
E-mail: K.A.Brookhuis@ppsw.rug.nl

Summary: Most traffic accidents may be attributed to driver impairment as a consequence of factors such as inattention, fatigue, intoxication, etc. It is now technically feasible to monitor and diagnose driver behaviour with respect to impairment with the aid of a limited number of in-vehicle sensors, however, a valid framework for the evaluation of driver impairment is still lacking. In order to provide an acceptable definition of driver impairment, a method is needed to assess both absolute and relative criteria. The seemingly paradoxical goal is to develop a definition of impaired driving that is consistent, yet adaptable to inter-individual differences.

INTRODUCTION

There is broad agreement among transport and traffic professionals on the one hand, and the driving community on the other, about what constitutes really unsafe or undesirable driving. The consensus is that sleepy, intoxicated or inattentive drivers constitute a safety hazard to themselves and others. Driver impairment due to such factors is sufficiently established to forecast a considerable decrease in accidents if electronic driver support systems were developed with the specific aim of detecting and attenuating impaired behaviour (Brookhuis et al., 2003a). Before this can be realised, however, the diagnosis of impaired driving behaviour must be established based on valid criteria (Brookhuis et al., 2003b), not on broad agreement.

There are a number of theoretical and practical problems surrounding the diagnosis of driver impairment beyond the anecdotal level. Most of the experimental research settles for an assessment of statistical significance, often at the alpha .05 level, which has little to say about the relevance of the established driving impairment effect in some arbitrary experiment by one or more factor levels. Ecological validity is not taken into account in the majority of cases.

In the simplest case, the diagnosis of impairment due to alcoholic intoxication is straightforward. The amount of alcohol consumed may be measured by Blood Alcohol Content (BAC), which has an exponential relationship with accident risk (Borkenstein et al., 1974), as shown in Figure 1. Based on this relationship, most countries have adopted a fixed level of BAC (mostly 0.5 promille) as a legal criterion. In other cases where the energetic state of the driver is inappropriate, such as under stress or fatigue, no equivalent index of impairment is available. However, a relationship similar to that in Figure 1 is likely (Brookhuis, 1998).
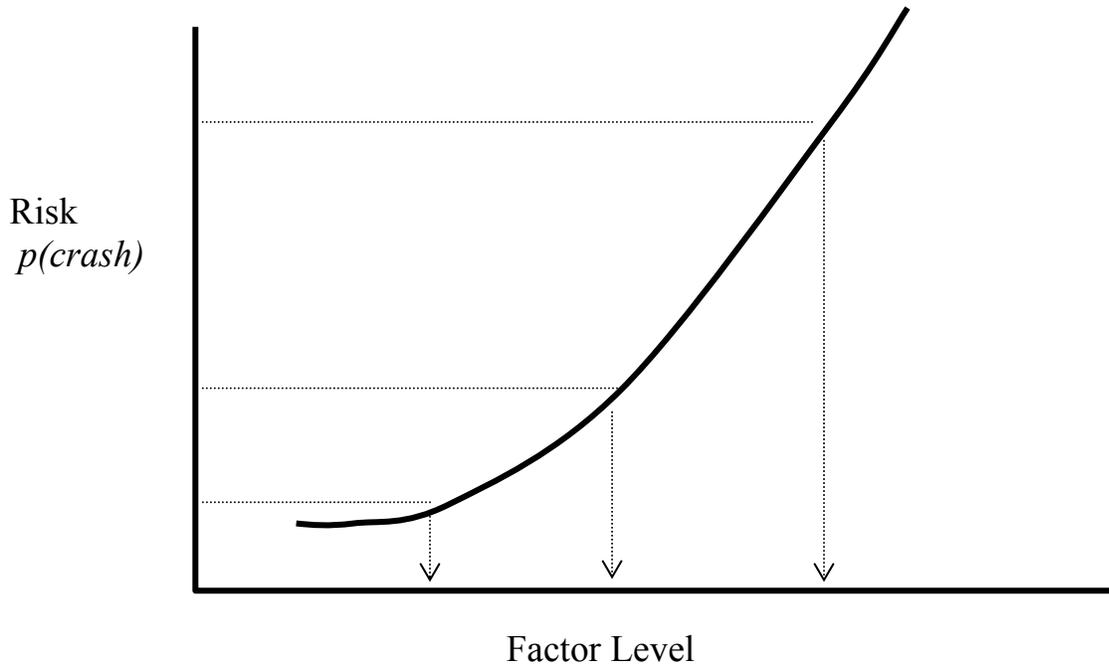
Figure 1. Typical, exponential relationship between the level of a factor (alcohol, licit and illicit drugs, fatigue, etc.) and accident likelihood.

This lack of an appropriate index may simply be a conceptual limitation. The individual often spontaneously perceives changes in energetic status in the absence of any explicit numeric scale or continuum. In the absence of an anchor scale (such as BAC in the case of alcohol, or body temperature in the case of impairment due to a thermal stressor), the multi-dimensional character of driver impairment renders the concept susceptible to an undesirable level of indeterminacy. This ambiguity at the conceptual level inevitably creates related problems of measurement and interpretation. This limitation is particularly striking when we attempt to measure the impact of multivariate energetic states on complex skills such as driving.

These problems set the background to the present analysis and discussion. There are practical reasons to devise a logical and consistent framework for the evaluation of driver behaviour. For example, it is technically feasible to monitor and diagnose driver performance with respect to accident-prone behaviour (Brookhuis & Brown, 1992; Brookhuis et al., 2003a) using real-time in-vehicle sensor apparatus. However, the feasibility of electronic sensor apparatus for this purpose is dependent on a valid framework for the evaluation of driver impairment (Fairclough et al., 1993; Brookhuis et al., 2003b). In other words, driver impairment should be defined first; i.e., criteria to decide on impaired driving performance should be assessed before an electronic driver behaviour monitor can be implemented. In broader terms, a large number of research papers are published each year in the field of traffic psychology, and a consensual framework to evaluate qualitative and quantitative aspects of driver impairment would aid comparisons between studies.

Definition of driver impairment

The first step towards a definition of driver impairment is to establish an accepted, valid distinction between "normative" and "impaired" categories of behaviour. One way to attain this goal is to induce impairment in a systematic and controlled fashion, thereby allowing us to study various levels on the continuum from normative driving to serious impairment. Somewhere along the continuum the criterion must be established for deciding if/when a driver is no longer able to perform adequately. This approach has been adopted in the European HASTE [1] project (Human Machine Interface And the Safety of Traffic in Europe), through the implementation of a series of secondary tasks of varying complexity and difficulty.

The technique is to manipulate the driving situation in such a way as to artificially induce impaired driving. The resulting data are then judged to be representative of impaired driving and may be suitable as a benchmark measure for comparison. One such straightforward method for inducing impairment is to have the driver consume a fair amount of alcohol, sufficient for a BAC above the legal level, and complete a journey in traffic. The detrimental effects of blood alcohol over the legal limit on traffic accidents are well documented, while the effects on driving performance have been demonstrated as well (Louwerens et al., 1987; Brookhuis, 1998). The data collected during a test ride may be used to benchmark impairment relative to normative driving for a particular driver. An alternative technique would be to expose the driver to such high levels of a stressor that the driver is unable to sustain safe performance and drifts off-road or collides with another vehicle. Extreme forms of imposed visual impairments, as introduced by De Waard et al. (1998), are feasible ways to induce stress in this sense. This approach has been used in a driving simulator by De Waard et al. (1998); they had subjects drive while they were distracted from the road by an extremely demanding visual task.

However, several alternative approaches are available as well. Since impaired driving usually implies that the driver is not fit to drive, psycho-physiological changes could be used as a classification index to define levels of driver alertness. For example, EEG data may be used to index impending sleepiness while carrying out tasks like driving, independently from behavioural measures (Brookhuis et al., 1986; Åkerstedt et al., 1991). It is possible to collect concurrent data from driving behaviour and some well-defined psycho-physiological measures, and to use the latter in order to categorise the former (Brookhuis, 1998). Among those are EEG as mentioned, but also heart rate and heart rate variability for effort and (mental) workload (Mulder, 1992; Mulder et al., 2003), facial muscle activity for effort and emotional load (Van Boxtel & Jessurun, 1993), and eyelid position and activity for sleepiness (Mallis & Dinges, 2003).

An alternative to the independent psycho-physiological measures would be expert rating. Professionals make an assessment of the individual's capability to drive safely based on their expert examination. Formalisation of expert observation may be used in order to develop a categorisation for safe versus impaired driving. A related, worked-out example is the use of observation of drivers' facial activity in order to index fatigue (e.g., Wierwille & Ellsworth,

---

[1] More information about the HASTE project can be found in Natasha Merat's contribution to this Proceedings, and derived from Karel Brookhuis or the co-ordinator Oliver Carsten (ocarsten@its.leeds.ac.uk).

1994). In this example, a system of scoring facial symptoms of fatigue was designed and observers were "trained" to score directly from videotape.

Feasibility and limitation of detecting impaired driving performance

All these techniques are plausible and feasible to set up a valid categorisation, but none are without limitations. In the case of *inducing impairment*, there is an implicit assumption that all categories of driver impairment are equivalent. Hence, impairment due to alcohol is treated as interchangeable with other variables such as sleepiness. The advantage of using alcohol in these conditions lies in its propaganda strength as a categorisation system based on legal and well-accepted criteria.

The technique described as *inducing impairment* provides a direct link between either behavioural or vehicle input measures and examples of impaired driving. Whilst ethical considerations dictate that the techniques described in the first part, by complex secondary tasks, are usually the domain of closed circuit or simulator investigations, those in the second technique, extreme diversion of driver's attention, must always be performed in the simulator. This may raise in principle valid questions regarding the perception of risk and drivers' motivation to avoid risk in a simulator. A second problem is the absence of any rationale when one chooses what size of time window is appropriate to define the period prior to a critical incident, e.g., a small window of 1 minute or a larger window of 10 minutes. The use of a regression-type analysis may also be questionable as time-on-task may affect two concurrent measures of driver impairment (e.g., EEG and lane keeping) and therefore create the impression that the two are directly related to one another.

Both *psychophysiology* and *expert observations* take the emphasis from primary task measures to covert indicators of energetic state that do not directly reflect the quality of driver performance. The advantage of these measures is an assumption of increased sensitivity. In addition, psycho-physiological measures are traditionally used to classify human sleepiness and therefore, the same approach may be applied to those stages of reduced alertness that occur prior to full-blown sleepiness. The weakness is that both these techniques represent an indirect index of driver impairment.

Finally, there is a problem concerning the consequences of signal-detection theory. In particular d-prime—i.e., the sensitivity of the impairment detection system—is relevant. Suppose the aim is to let the system discriminate drunk driving from normal, sober driving by a simple, easily derived driving parameter, the amount of weaving, measured as the standard deviation of lateral position (SDLP). Taking the case of alcohol consumption has the inherent advantage of representing a benchmark, i.e., the relationship between BAC and accident likelihood (Borkenstein et al., 1964). Based on this relationship, in some countries the legal limit of BAC is 0.5 promille, in others it is still 0.8 promille. From the data of Louwerens et al. (1987), average SDLPs with these two BACs are known values and could thereupon be taken as criteria on the earlier mentioned continuum to decide about the boundary between normative or safe and impaired driving performance. However, the distributions of the curves under normative and impaired driving overlap in this case (see Brookhuis et al., 2003b). Avoiding false alarms, which is vital with respect to acceptance (and thus use) of this type of driver performance monitoring system, implies serious restrictions.

CONCLUSION

An adequate representation of driver behaviour is fundamental for the development of criteria and categorisation of impairment. These criteria are formulated to define the division, or "red line" between the normative and impaired examples of the primary task. These criteria may be formulated in either absolute or relative terms. The former relates to absolute values of behavioural measures, valid under all circumstances; the latter relates to individual differences. For an extensive explanation of the difference between absolute and relative criteria, see Brookhuis et al. (2003b).

REFERENCES

Åkerstedt, T., Kecklund, G., Sigurdsson, K., Anderzén, I., & Gillberg, M. (1991). Methodological aspects on ambulatory monitoring of sleepiness. In *Proceedings of the workshop Psychophysiological Measures in Transport Operations*. Köln: DLR, pp.1-21.

Borkenstein, R.F., Crowther, R.F., Shumate, R.P., Ziel, W.B., Zylman, R. (1964). The role of the drinking driver in traffic accidents. *Blutalkohol, 11*, suppl. 1.

Brookhuis, K.A., Louwerens, J.W., & O'Hanlon, J.F. (1986). EEG energy-density spectra and driving performance under the influence of some antidepressant drugs. In J.F. O'Hanlon & J.J. de Gier (Eds.), *Drugs and Driving*, (pp.213-221). London: Taylor & Francis.

Brookhuis, K.A., Brown, I.D. (1992). Ergonomics and road safety. *Impact of science on society, 165*, 35-40.

Brookhuis, K.A. (1998). How to measure driving ability under the influence of alcohol and drugs, and why. *Human Psychopharmacology, 13,* 64-69.

Brookhuis, K.A., De Waard, D., Kraaij, J.H., Bekiaris, E. (2003a). How important is driver fatigue and what can we do about it? In: D. de Waard, K.A. Brookhuis, S.M. Sommer, W.B. Verwey (Eds.) *Human Factors in the Age of Virtual Reality.* Maastricht: Shaker Publishing, 191-207.

Brookhuis, K.A., De Waard, D., Fairclough, S.H. (2003b). Criteria for driver impairment. *Ergonomics, 46,* 433-445.

De Waard, D., Van der Hulst, M., & Brookhuis, K.A. (1998). The detection of driver inattention and breakdown. In: P. Albuquerque, J.A. Santos, C. Rodrigues, A.H. Pires da Costa (Eds.), *Human Factors in Road Traffic II.* (pp. 102-108). Braga: University of Minho.

Fairclough, S.H., Brookhuis, K.A., Vallet, M. (1993). Driver state monitoring system DETER (V2009) (pp. 330-335). In *Advanced Transport Telematics, Proceedings of the Technical Days.* Brussels: Commission of the European Communities.

Louwerens, J.W., Gloerich, A.B.M., De Vries, G., Brookhuis, K.A., & O'Hanlon, J.F. (1987). The relationship between drivers' blood alcohol concentration (BAC) and actual driving performance during high speed travel. In: P.C. Noordzij and R. Roszbach (Eds.), *Alcohol, Drugs and Driving - T'86*, (pp.183-187). Amsterdam: Excerpta Medica.

Mallis, M.M., Dinges, D.F. (2003). Monitoring Alertness by Eyelid Closure. In: N. Stanton, A. Hedge, H.W. Hendrick, K.A. Brookhuis, E. Salas (Eds.). *Handbook of Ergonomics and Human Factors Methods*. London: Taylor & Francis.

Mulder, L.J.M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology, 34*, 205-236.

Mulder L.J.M., De Waard D., Brookhuis K.A. (2003). Estimating Mental Effort using Heart Rate and Heart Rate Variability. In: N. Stanton, A. Hedge, H.W. Hendrick, K.A. Brookhuis, E. Salas (Eds.). *Handbook of Ergonomics and Human Factors Methods*. London: Taylor & Francis.

Van Boxtel, A., & Jessurun, M. (1993). Amplitude and bilateral coherency of facial and jaw-elevator EMG activity as an index of effort during a two-choice serial reaction task. *Psychophysiology, 30*, 589-604.

Wierwille, W.W. & Ellsworth, L.A. (1994). Evaluation of driver drowsiness by trained raters. *Accident Analysis and Prevention, 26,* 571-581.