ORIGINAL PAPER

CrossMark

The use of neuropsychological tests to study the effects of aging on driving performance in the UK

Rawid Khan[1] · Muhammad Tariq Khan[2] · Bashir Alam[1]

Abstract

Research was conducted to identify a series of neuropsychological tests to assess the ability to drive. Driving performance of young and old UK drivers was modeled through multiple linear regression and univariate logistic regression tools. The UFOV3 test (i.e. the third subtest of the UFOV test) had comparatively high discriminating ability in separating poor-drivers from not-poor-drivers, with 92.86% of the drivers correctly classified; the UFOV3 test resulted in a Sensitivity of 62.5%. Age and a composite cognitive measure were also found to be sound discriminators of poor-drivers and not-poor-drivers with 91.07% and 89.28% of the drivers correctly classified respectively; both age and the composite cognitive measure resulted in a Sensitivity of 50%. It was found that the commonly recommended Clock Drawing Test and the Trail Making-B test were insignificant predictors of driving ability. Results suggest that for a score greater than 220 on the UFOV3 test, the driver may be further evaluated by a driving specialist to ascertain questionable driving behavior. Also, drivers above the age of 77 were more susceptible to exhibiting unusual driving behavior; if such drivers have UFOV3 scores greater than 220 it would be more appropriate to evaluate driving behavior through a driving specialist.

Keywords Mild cognitive impairment · Neuropsychological tests · UFOV test · Dichotic listening test · Regression analysis

1 Introduction

The proportion of licensed drivers is increasing in the general driving population and a substantial number within this population group are experiencing a neuropsychological decline in functions that are critical to the driving task. According to one estimate, about 40% of the driving population will be over the age of 60 by the year 2020 in the UK and currently, several hundred thousand drivers with dementia hold driving licenses [15]. In police reported crashes, older drivers are 2.9 times more likely to die than middle aged drivers [9]. Researchers point out that since older drivers are susceptible to easily get injured and accidents are reported on the basis of personal injury or fatality, this leads to a sampling bias which shows an increase in age

related risk for older drivers according to official statistics [18]. Using the decomposition method of analysis, it was shown that sustained serious injury rates of the younger driver group (30 to 50 years) was less than half that of the older driver group (above the age of 70) [30]. On the contrary, young drivers' crash rates surpass those of older drivers when evaluated based on crashes per kilometer driven [22]. It was also pointed out that the higher casualty rate per mile reported for the older driver group was on account of the fact that as drivers get older they tend to have lower annual mileage driven [4]. According to the OECD (Organization for Economic Cooperation and Development) working group, older drivers do not present an unacceptable risk to other drivers/pedestrians but rather are more prone to be self-injured in accidents [44]. Researchers while studying two-vehicle crashes for 1992–1994 as recorded by police found that when risk estimates for elderly drivers are based on the number of licensed drivers, they do not constitute an accurate analysis of the issue [11]. Furthermore, utilizing US data from 1994 to 1996 relevant to crashes with regard to age and sex of road users, it was found that when seventy-years-old male drivers had their annual license renewal, on average they posed 40% less of a threat to other drivers/pedestrians than the annual license renewal of forty-year-old male drivers [13]. Older healthy drivers

✉ Rawid Khan
  dr.rawid@uetpeshawar.edu.pk

1 Department of Civil Engineering, University of Engineering and Technology, Peshawar, Pakistan

2 Department of Civil Engineering, University of Southampton, Southampton, UK

 Springer

perform the driving task at a level that is comparable with fit young adults. Since dementing illnesses are common in old age, a certain proportion of older drivers are in the early stages of a dementing illness or already clinically demented.

Since the neuropsychological decrement associated with normal ageing cannot be readily distinguished from that of very early stage dementia, and it is very difficult to diagnose the disease in the early stage; a considerable number of older drivers may continue driving because many of them will not be diagnosed as having the disease by physicians [6, 33, 40]. Brown and Ott [5] report that there is evidence to support that not all persons in the early stages of dementia are incompetent drivers. In fact, the performance of drivers with mild neuropsychological impairment is not consistently significantly worse than their healthy counterparts on the on-road test [35]. For individuals that have moderate to severe dementia, there is strong consensus that they should not drive, however decisions regarding those having mild dementia are problematic [7]. Therefore, some individuals who have mild dementia, possess sufficient driving skills to be designated as fit drivers; however, a stage/time will come when their neuropsychological impairment will exacerbate and will ultimately render them unfit drivers. It was ascertained that all European countries do not conduct cognitive screening of older drivers and the advantages of driver screening based on age do not outweigh its dis-benefits [38]. This asserts the necessity of neuropsychological testing in context of driver screening.

The most challenging assessment and decision for the licensing authority/physician as regards fitness to drive lies in drivers who are questionably demented or are in a state of very mild dementia. In the United Kingdom, more emphasis is placed on self-declaration of illness (to the DVLA) i.e., license holders are required under the law to inform the DVLA (Driver and Vehicle Licensing Agency) of any health condition that may influence their ability to drive. Subsequently, it is the responsibility of the DVLA Medical Branch to make judgment regarding a person's fitness to drive. Before making a judgment, the DVLA may adopt the following course of actions [10]: (a) request additional information from the GP/Consultant, (b) referral for specialist clinical assessment, or (c) acquiring an independent medical opinion. Some times as the DVLA deems necessary, it may require individuals to re-take the standard driving test or may refer him or her to a specialist driving assessment centre. Sometimes the family doctor/GP on his own initiative may contact the Medical Advisory Branch of the DVLA.

While carrying out a meta-analysis of 27 studies of drivers with dementia and their neurophysiological ability, there was wide variance among the findings [37]. The development of a particular selection of neurophysiological tests for the screening of drivers has been prevented by hurdles like unawareness of specific neurophysiological domains that are relevant to driver behavior [29]. Also, sound neurophysiological testing

protocols and service policies are lacking which would allow the integration of neurophysiological, driving evaluation data and medical data in order to facilitate DVLA and drivers make decisions regarding fitness to drive [10]. While evaluating neurophysiological status in primary care settings, researchers through survey data established that physicians exhibited a lack of confidence in assessment of drivers [31]. There is no single neuropsychological test that can reliably and economically separate safe older drivers from those that are distinctly unsafe by identifying all deficits that are crucial to driving. Even the on-road driving tests may not identify important driving deficits in older drivers. Hence there is no standard testing protocol (that is reliable) for assessing a person's fitness to drive after the onset of neurological disease/trauma and/or natural ageing. Therefore, different neuropsychological tests tapping different cognitive domains are in use. Hence, in the absence of a standard reliable protocol, the decisions regarding fitness to drive, are doubtful and exude a low level of confidence on part of the clinicians/professionals. Due to the lack of a reliable standard protocol, some clinicians make their judgments based on self-report (of drivers), which has risks associated with it as lack of insight and judgment are potential common traits of the population experiencing neuropsychological decrements. In this context, Christie et al. [10] while carrying out a survey of clinical psychologists with regard to neuropsychological settings in the UK in assessing fitness to drive after head injury, observed: "Overall, clinicians' decisions about a client's fitness to drive seem to be based on an eclectic approach with considerable reliance on *clinical impression*". Seldom is recourse made by health professionals to driving assessment as a first alternative as it requires a fee and such centres are not readily available everywhere. Thus there exists a need for more information on assessment of fitness to drive with regard to neuropsychological tests, since medical information alone is not sufficient to assist in decision making of fitness to drive. This will also alleviate the need for the requirement of an on-road evaluation/assessment or can be a supplementary tool in addition to on-road assessment and will instill more confidence in decision making on part of the clinician. The Driver and Vehicle Licensing Agency (DVLA) of the UK publishes and updates a guide for medical professionals with a view to assisting them in assessing fitness to drive [12].

The objective of this research was the identification of a series of neuropsychological tests suitable to be used to assess an individual's ability to drive and especially of those individuals that are questionably demented or are in a subtle state of cognitive decline (and therefore are the most difficult to identify). The relative ability of the neuropsychological tests to discriminate between "poor-driving" and "not-poor-driving" and the effect of other factors such as age and a composite cognitive measure (based on all nine neuropsychological tests) was also evaluated in discriminating "poor-drivers" and "not-poor-drivers".

2 Methodology

A schematic diagram of the methodology adopted in carrying out this research is shown in Fig. 1. Since drivers crash more frequently in a simulator than in real life [28], it was decided to use a comparative *approach* to evaluate driving performance (or decline in driving competence) by testing experienced younger drivers and experienced older drivers. The selection of a random sample of drivers was neither feasible nor desired. Therefore, a convenient sample of volunteers (from both age groups) was sought. The selection bias introduced as a result of the non-random sample was not of much concern as it was not intended to generalize the estimated percentage of drivers (with impaired driving ability etc.) to the entire driving population i.e., we were not interested in estimating the proportion of drivers (with decrements in driving ability) in the population but rather were interested in the range of performance capabilities of drivers. Also, there were good prospects that the non-random sample may enable the testing of some "information rich" cases of older drivers which were really crucial and valuable.

For the younger age group, the statistically safest age group of 26 to 40 years was selected whereas for the older group the age limit was fixed as above 60 years [16, 17]. Researchers studied older driver in a driving simulator of ages 55 and older [2]. Younger-onset dementia (early-onset dementia) occurs before the age of 65 [19, 20]. In this study, the older group of drivers was of ages 60.3 to 88.4 years (Table 1). All drivers were to have a valid UK driving license and be current drivers with at least 5 years of driving experience. Also, all drivers were to have adequate visual, hearing, communication and physical capabilities to complete the simulator driving tests/

Fig. 1 Schematic diagram showing the methodology of research

assessment. In total, fifty six drivers (28 from each group) were successfully tested. The demographic detail of the *successfully-tested* subjects may be seen in Table 1.

Approval for the research was obtained from the School Research Ethics Committee. Advertisement leaflets were distributed in the Southampton area and in a number of bowling clubs. All subjects were tested in the morning so as to avoid systematic effects of fatigue. The participants were first given a short (3 to 4 min) run on the practice drive (the beginning portion had S-curves, which can expose drivers prone to simulation sickness) in the driving simulator to ensure that the driver was not prone to simulation sickness syndrome. Nausea, disorientation and ocular problems such as eyestrain, blurred vision and eye fatigue have been reported as some of the indicators of simulation sickness in fixed-base simulators [27, 32]. If a participant experienced the syndrome, the practice drive was immediately terminated and the participant was deemed unfit to take the simulation drive. If a participant did not feel any discomfort in the practice drive, then the rest of the protocol followed. The sample of drivers was tested over a period of 3 months. The annual road mileage of participants was not considered as most drivers (especially older drivers) confessed to its inaccuracy. This inaccurate reporting of annual mileage by older drivers may have been consciously reported because of its roots in the low-mileage bias phenomena [26].

Participants were given the following neuropsychological tests in random order: (1) UFOV Test (consisting of UFOV1, UFOV2 and UFOV3) (2) Dichotic Listening Test (3) Trail-Making Test (4) Rey-Osterrieth Test (5) Paper Folding Test (6) Clock Drawing Test (Freedman scoring algorithm). This was followed by a practice drive for the Main Drive (Drive-I) and then the Main Drive (Drive-I). This procedure was repeated for the DA and Car-Following Drive (Drive-II). Frequent breaks for refreshments were provided (but not in the middle of a simulation run or a neuropsychological test).

Drive-I consisted of a 21 mile drive (duration approximately 40 min). Some of the elements incorporated into the drive were: (a) Controlled Hazards (pedestrian/dog intrusions, intersection intrusions, sudden-braking, intrusions with limited sight distance etc.) (b) Right turns across oncoming traffic (c) Left turns (involving cyclist) (d) Dangerous overtaking by opposing vehicle (e) Lane changing manoeuvres (in traffic) and lane drops (f) Stop controlled intersections (gap selection) (g) Overtaking manoeuvres in the wake of a stream of oncoming vehicles (h) Transiting construction zones (i) Signalized intersections (j) Signalized intersections with dilemma zones. (k) Gas station manoeuvres. (l) Tracking task (boxes fallen from truck on the road, mountain S-curves). The drive was predominantly rural with an urban flavour in certain reaches. Infrastructure consisting of static objects such as buildings, parked cars, trees, and road signs etc. was added and was thoughtfully designed giving due regard to as to how they would affect a driver's visual search pattern, perception,
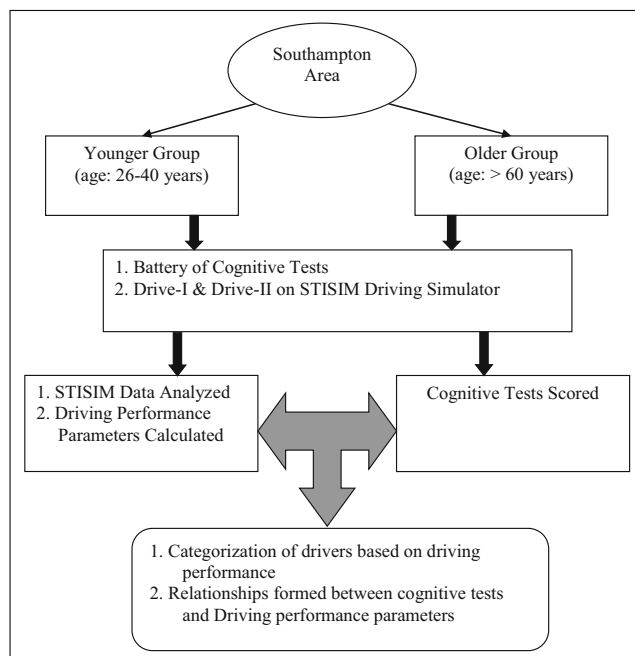
Table 1   Demographic detail of the successfully-tested younger and older driver groups

Group	No. of Males	No. of Females	No. of Total Subjects	Minimum Age (yrs)	Maximum Age (yrs)	Mean Age (yrs)	Median Age (yrs)	Standard Deviation of Age (yrs)
Young	14	14	28	26.3	40	32.3	32.3	4.4
Old	16	12	28	60.3	88.4	68.7	66.2	7.4

attention and/or driving behaviour in context of the critical event. Telephone poles at a spacing of 200 ft were placed along the road alignment to give drivers cues as to their speed.

The DA (Divided Attention) and the Car-Following drive collectively designated as Drive-II, was a 14 mile drive with a total duration of 16 min, with about 8 min consumed by the DA portion. The first portion of the drive comprised of a DA task while the second portion was a Car-Following task. In order to assess continuous measures of driving ability (e.g., speed, steering control/lane keeping), the driver or the vehicle can be stimulated in a controlled manner e.g., by the application of wind gust, lead vehicle with a controlled velocity profile or a divided attention task etc., and the driver's response to the stimulus measured [1]. Due to the workload from competing sources, it is not possible for the driver to respond in an optimum manner to the primary task (driving) and secondary task (e.g., DA task) and one or both are bound to suffer. This trade-off can be measured and may show up as unfavourable performance on one or more driving performance parameters such as an increased reaction time, deterioration in lane positioning or speed adherence [23].

The differences between driving performance may also reflect the ability to drive in a simulator; in order to minimize this effect, we included (a) a practice drive on the simulator and only when the participants felt comfortable and at ease with the simulator did we proceeded further with the full test in the simulator, (b) neither participants in either group (old and young) had previous experience of driving in a simulator, thereby minimizing the confounding effect of practice.

Note: The UFOV3 test is the third subtest of the UFOV (Useful Field of View) test and takes up the most cognitive effort compared to UFOV1 and UFOV2. This test cannot be administered in isolation, because the examinee has to get gradually accustomed to the difficulty of the task by first familiarizing himself with UFOV1 and UFOV2. The UFOV test takes only 15 min to administer using an ordinary on-the-market personal computer with a 17 in. monitor and does not require computer literacy

2.1 Simulation

Usually, the normal on-road test is insufficiently challenging to identify risky driving behavior due to cognitive impairment [34]. Therefore, it was decided to address the research objectives through the application of Driving Simulation. Also it has

been found that there is significant correlation between the on-road driving index and the driving simulator [8]. A simulation drive was programmed using the STISIM® software [41].

Previous studies have categorized drivers as having "failed" when their driving performance score (as gauged by some kind of a driving performance index/score) fell more than 2 standard deviations below the mean of the control group (the normal group); this criterion ensures that, in a normally distributed population, approximately 2.3% (probability $Z \leq -2$ is 0.023) of the normal drivers will always be considered as abnormal even if their performance does not qualify for a "poor/failed" status; besides, this criterion cannot be used for categorization of drivers if more than one driving performance measure is used simultaneously to assess driver performance (which is crucial). In fact in a study [8], the authors have acknowledged the non-discriminatory nature of using a single "driver error" variable i.e. a single driving performance index.

In order to identify especially those drivers that are the most challenging to identify (because they are questionably demented or are in a state of very mild dementia), a methodology was developed by which drivers were categorized by considering more than one (i.e. three) driving performance indices/parameters simultaneously and a "poor driver" group identified; detailed methodology is given in the author's thesis [23]. This methodology entailed: (1) testing a diverse sample of subjects (non-clinical sample of Table 1), (2) simulation drive-designs (Drive-I and Drive-II) based on specific psychometric principles wherein 24 driving performance parameters were monitored, (3) calculation of driving performance indices by removing parameters contributing to "noise" and keeping the ones contributing to "signal" through the concept of Cronbach's Alpha Reliability Coefficient and weighting, and (4) the technique of Normal-Mixture-Model Cluster Analysis.

Three different kinds of unit nominal weight indices were computed. These were: (1) Index obtained by considering all 24 driving performance parameters (Index named = DPI1). (2) Index obtained by considering all 24 driving performance parameters except No. of Total Hazards (Index named = DPI2). (3) Index obtained by considering all 24 driving performance parameters except No. of Total Hazards and No. of Low-Speed Warnings (Index named = DPI3).

Weighted versions of the indices DPI1, DPI2 and DPI3 were called DPI1-weighted, DPI2-weighted and DPI3-weighted respectively and were arrived at by the multivariate statistical

Table 2 Mean values of DPI3-weighted, No. of Total Hazards and No. of Low-speed Warnings for model EEV with BIC of −1075.63

Groups for model EEV* with 3 clusters/groups having BIC of −1075.63

Group no.	Number of drivers	Mean No. of total hazards	Mean No. of low-speed warnings	Mean DPI3-weighted
1	20	8.10	441.55	−0.26
2	8	16.50	553	−0.90
3	28	7.32	381.04	0.44

Maximum & Minimum value of No. of Total Hazards = 30, 2

Maximum & Minimum value of No. of Low-Speed warnings = 656, 286

Maximum & Minimum value of DPI3-weighted = 1.16, −2.08

BIC, Bayesian Information Criterion

*= EEV is a specific Model selected in cluster analysis

technique of Principal Component Analysis. Principal Component Analysis was used to find differential weights so as to maximize Alpha. The use of differential weights was used to maximize reliability (Alpha). More reliability translates to more variance of the composite score (i.e., more dispersion of composite score) and therefore yields more discrimination among individuals [24]. The first Principal Component of the standardized variables (parameters) maximizes the explained variance and therefore, its eigenvector furnishes the weights that maximize Cronbach's Alpha. Using the technique of Normal-Mixture-Model Cluster Analysis [14] where six models/scenario were used resulted in the best model which included the simultaneous use of the three variables DPI3-weighted, No. of Total Hazards and No. of Low-speed Warnings.

This approach resulted in the identification of a "poor driver" group which even included some individuals exhibiting relatively subtle changes in their driving performance. Table 2 shows the group means of the three cluster/group (i.e. three categories of drivers) model finally selected on account of its higher BIC value in cluster analysis that was based on the driving performance parameters/index No. of Total Hazards, No. of Low-Speed warnings and DPI3-weighted. The driving performance parameter "No. of Total Hazards" was the sum of No. of Off-road Accidents, No. of Collisions, No. of Pedestrian Hits, No. of Traffic Light Tickets, No. of Stop Signs Missed, No. of Illegal Turns and No. of Stops in Middle of Traffic. All these parameters are discrete events that represent a substantial risk of crashes and traffic conflicts and they signal declining driving skill. The driving performance parameter "No. of Low-Speed Warnings" enumerated the number of "ding" sounds played every three seconds when the driver's speed was more than 5 mph below the posted speed limit. Overcautiousness (e.g., driving slowly) has been classified as a discriminating error [35, 39, 43]. Discriminating errors are potentially dangerous errors that signify degradation in driving skill [23].

Groups of drivers (driving proficiency wise) identified through cluster analysis shown in Table 2 indicate that on

average, group no. 3 had the highest scores on the index DPI3-weighted, the lowest No. of Total Hazards and the lowest No. of Low-speed Warnings compared with group no. 1 and 2. Also, group no.1 had higher score on the index DPI3-weighted, lower No. of Total Hazards and lower No. of Low-speed Warnings compared with group no.2. Therefore, driving performance in decreasing order of skill by group number is: group no.3, no.1 and no. 2. Despite the fact that drivers in Group no.2 were driving on average at the lowest speeds, they had the greatest number of accidents etc. (i.e. no. of total hazards) and had low rating on all other driving performance measures (i.e. DPI3-weighted). Group no.2 was the smallest group (comprising of 8 drivers all from the older driver group) and may be considered as possessing poor driving skills. Table 3 shows neuropsychological performance on the nine neuropsychological tests on average was the highest by group no.3, then group no. 1 and then group no. 2. This same order was also observed in decreasing order of driving-performance-skill among the three driver groups (Table 2).

Figure 2 shows a driver-classification graph of No. of Low-speed Warnings against No. of Total Hazards for the driver groupings of Table 2; the ellipses superimposed on the classification plot in Fig. 2 correspond to the multivariate analogs of the standard deviations for each mixture component (i.e. they correspond to the covariances of the components) with centers at the means μ_k. Table 2 and Fig. 2 indicate that the attributes of group No. 2 were quite deviant from the other two groups (i.e. group No. 1 & 3), which were relatively quite close. It was logical to merge group No. 1 and 3 to form a "not-poor-drivers" group and designate group No. 2 as the "poor-drivers" group. Thus the "poor drivers" group consisted of 8 drivers and the "not-poor-drivers" group of 48.

2.2 Neuropsychological tests

It is possible to predict fitness to drive through neuropsychological tests or psychiatric measures. Multiple cognitive domains are called upon when an individual is negotiating

Table 3 Average scores on Neuropsychological Tests

Group No.	Trail[*]	Clock	Rey-Copy	Rey-Recall	Dichotic*	Paper	UFOV1*	UFOV2*	UFOV3*
1	81.75	14.80	32.05	19.45	1.85	38.47	18.90	31.30	113.45
2	114.87	14	29.37	15.93	7	24.75	20.12	86.75	241
3	64.78	14.92	33.50	23.82	1.41	52.01	17.10	19.39	75.64

*Higher score corresponds to lower neuropsychological performance on the Neuropsychological Test

driving scenarios/situations [35]. We must bear in mind that many of the cognitive constructs themselves are interrelated [3] and interact with each other [2]. Nine neuropsychological tests were identified to cover key cognitive domains necessary for safe driving. These tests assess a broad range of neuropsychological skills and are in the public domain with the exception of one test (i.e. UFOV test). Significant diversity with regard to administration of these tests is exhibited as they include paper and pencil tests, listening test and visual computer controlled test. Also, these tests are quite sensitive to the effects of ageing and to a range of diseases that are well known to impair driving performance. With regard to neuropsychological tests we must bear in mind that a single test does not reflect a pure measure of a single cognitive domain, but rather each test taps more than one cognitive domain. Also, each test only partially taps a specific domain. Therefore, keeping these key points in view, we selected more than one test that tapped the same domain that is very critical relevant to the driving task e.g., there was more than one test that tapped visuospatial abilities and attention (because these are highly crucial domains relevant to driving) [35]. Relevant to the neuropsychological tests, detailed literature review and instructions (adopted) can be found in Khan [23].
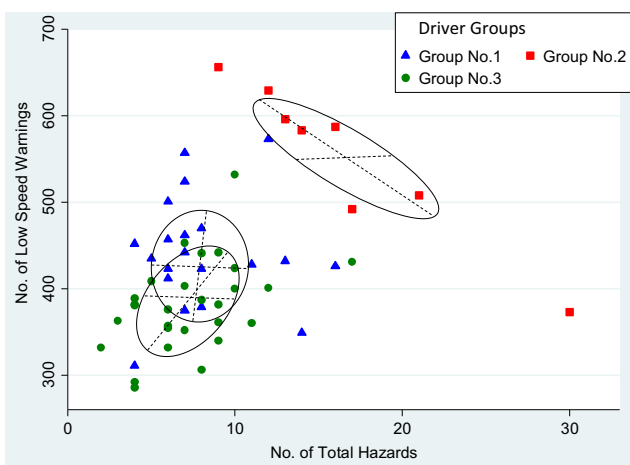

Fig. 2 Driver-classification graph of No. of Low-speed Warnings against No. of Total Hazards for model EEV with 3 groups/clusters having BIC of −1075.63

2.3 Linear and logistic regression models

Parsimonious multiple linear regression models were developed using the nine neuropsychological tests as predictors in order to predict general driving ability. To model (linear regression model) the driving performance indices through the different neuropsychological tests it was necessary that the indices included the effects of all viable driving performance parameters i.e. the indices that were derived by considering all 24 driving performance parameters, which were DPI1 and DPI1-weighted. Since the Pearson Moment Correlation coefficient between DPI1 and DPI1-weighted was very high (0.9975), it was decided to model DPI1-weighted (dependent variable) through neuropsychological tests (independent variables), as this index was geared to provide maximum discrimination between drivers because of Principal Component weights.

Logistic regression models were developed in order to discriminate "poor-drivers" from "not-poor-drivers" by using: (1) A single neuropsychological test amongst the nine that brought about the best discrimination (2) Age as predictor (3) A composite measure of all nine neuropsychological tests as a predictor. To model the "poor-drivers" through logistic regression using different neuropsychological tests, it was necessary that the dependent variable (driver grouping) be dichotomized into "poor drivers" and "not-poor-drivers" from the three-group classification of Table 2. Table 2 and Fig. 2 indicate that the attributes of group No. 2 are quite deviant from the other two groups (i.e. group No. 1 & 3), which are relatively quite close. Thus the "poor drivers" group consisted of 8 drivers and the "not-poor-drivers" group of 48. Following the conventions of logistic regression and our objective, the 8 drivers were coded as 1 (i.e. the event of interest occurred) and the remaining 48 as zero (i.e. the event of interest did not occur) under the dependent variable category. Keeping in view the overall sample size, the number of candidate predictors, and the lop sidedness of the dependent variable (i.e. 8 versus 48), the decision to develop a multivariate logistic model based on all possible predictors (trail, clock, rey-copy, rey-recall, dichotic, paper, UFOV1, UFOV2, UFOV3) or a smaller subset of them was not appropriate as it would have resulted in a numerically unstable model (thereby making use

of the likelihood ratio tests unreliable). Instead, as per recommendation of Hosmer et al. [21] it was decided to fit nine univariate logistic regression models, one for each predictor (i.e. each neuropsychological test) and use a variety of Goodness-of-fit statistics/criterion such as Deviance, BIC (Bayesian information criterion), AIC (Akaike's information criterion), Pseudo-R^2 and p-values (likelihood ratio test) to aid decision making in selecting the best model.

3 Analysis and results

The best/parsimonious model developed was:

$$DPI = 104.31 + 23.40(dichotic + 1)^{-0.5} - 7.83 \log_e(ufov3) + 0.72(rey\text{-}recall) \tag{1}$$

This model developed with dichotic, UFOV3 and rey-recall as independent variables (equation 1) had the lowest BIC value (267.95) and a higher R^2 value (0.59). The Driving Performance Index (DPI) in this model may be regarded as a general driving performance index that was obtained by considering all 24 driving performance parameters. As a guide for prediction, predicted scores less 96 (i.e. 2 standard deviations below the mean of the younger group) would signify poor driving proficiency and greater than 110 (median of the younger group) would signify good driving proficiency. This index gives an idea of the general driving skill of a driver with essentially the same emphasis being placed on each driving performance parameter, be it for example the No. of Total Hazards, Over Speed Limit (Percent of Time) or Standard Deviation in Speed DA Task and therefore, cannot be used in the identification of drivers exhibiting risky driving behavior due to neuropsychological impairment. In order to identify drivers exhibiting risky driving behavior due to neuropsychological impairment, it was necessary that the effects of parameters that assess driving skills at the "controlled processing level" ("effortful" processing) be isolated from the rest of the parameters, before undertaking evaluation of risky-driving behavior based on these broader categories of parameters. Since these effects were not isolated in the Driving Performance Index (DPI) in the model, therefore this index is not capable of discriminating between risky driving behavior due to neuropsychological impairment and normal driving behaviour.

The best logistic regression model was the univariate model based on the UFOV3 neuropsychological test because of the lowest BIC, AIC and even Deviance. The Pseudo-R^2 of this model was also the highest and the P-value (0.0001) for the likelihood ratio test indicated that the model was statistically highly significant. The next best univariate models in decreasing order of overall fit were the models based on dichotic, trail, rey-copy and paper neuropsychological tests. The best logistic regression model (based on the UFOV3 test, i.e.

Table 4 Coefficients for univariate logistic regression using UFOV3, age and *Tscore_cognitive*

Predictor	Coefficient	Standard Error	z	P > |z|	95% Conf. Interval	
UFOV3	0.019135	0.0064765	2.95	0.003	0.0064412	0.0318288
age	0.1654765	0.0659499	2.51	0.012	0.0362171	0.2947359
Tscore_cognitive	0.1067349	0.0366649	2.91	0.004	0.034873	0.1785968

equation 2) was subjected to diagnostic checks/goodness-of-fit test and confirmed through model validation.

$$\hat{\pi} = \frac{1}{1 + e^{-(-4.591905 + 0.019135\ ufov3)}} \tag{2}$$

Where π is the probability of the outcome of interest (i.e. probability of being a "poor-driver"). An Odds Ratio (based on a unit increase in UFOV3) of 1.019319 and 2.15 (based on a 40 unit increase in UFOV3) was obtained. Tables 4 and 5 show standard errors and confidence intervals for coefficient of UFOV3 and odd ratio (based on unit increase in predictor) of the model. Tscore_cognitive is a composite neuropsychological measure of the nine neurophysiological tests.

This means that the odds of being a "poor-driver" increase over 2.15 fold for drivers having an increased (i.e. difference on the higher side) UFOV3 score of 40 compared with the UFOV3 scores of other drivers or in other words, the odds of being a "poor-driver" increase 115% for drivers having an increased (i.e. difference on the higher side) UFOV3 score of 40 compared with the UFOV3 scores of other drivers. An area under ROC curve (Receiver Operating Characteristic curve) of 0.8659 was obtained (Fig. 3) which is considered excellent discrimination [21]. According to Kutner et al. [25], selecting a cut-point value (of predicted probability) of 0.5 is only reasonable when (a) it is equally likely for the outcome of interest (i.e. "poor-drivers") and the complementary outcome (i.e. "not-poor-drivers") to occur in the population of interest; and (b) the cost of incorrectly predicting the outcome of interest (i.e. "poor-drivers") and the complementary outcome (i.e. "not-poor-drivers") are approximately the same. Since these two conditions were not satisfied therefore a cut-point of 0.5

Table 5 Odd Ratios for univariate logistic regression using UFOV3, age and *Tscore_cognitive*

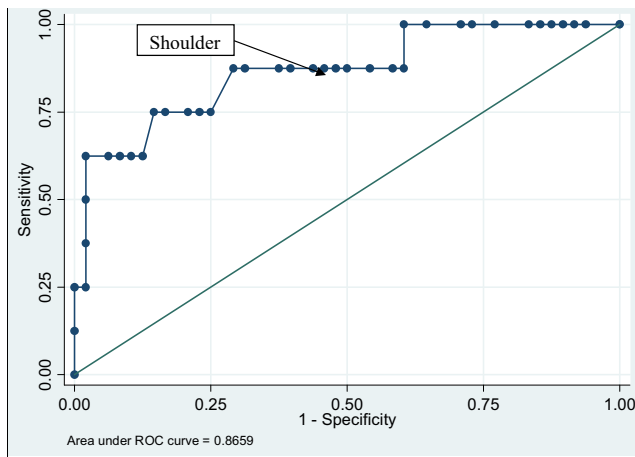| Predictor	Odds Ratio (based on unit increase in predictor)	Standard Error	z	P > |z|	95% Conf. Interval	
UFOV3	1.019319	0.0066017	2.95	0.003	1.006462	1.032341
age	1.179955	0.0778179	2.51	0.012	1.036881	1.342772
Tscore_cognitive	1.112639	0.0407948	2.91	0.004	1.035488	1.195539

Fig. 3 Receiver Operating Characteristic (ROC) curve for different cut-pints (predicted probabilities used as cut-points). Area under ROC curve is 0.8659

was avoided. As an alternative, Kutner et al. [25] suggest to use a cut-point value such that the proportion of incorrect predictions is lowest (or the proportion of correct predictions is highest). A cut-point of 0.4 provided the highest correct classification rate of drivers and resulted in the following figures: (i) Sensitivity = 62.5%, (ii) Specificity = 97.92%, (iii) positive predictive value = 83.33%, negative predictive value = 94%, correct classification = 92.86% (Table 6). The cut-point value of predicted probability of 0.4 when used in logistic regression Eq. 2 corresponds to a cut-point value of 219 (after rounding) of UFOV3. At this threshold, three "poor-drivers" and one "not-poor-driver"' were misclassified (i.e. false negatives = 3, false positives = 1). 92.86% of the drivers were correctly classified as either being "poor-drivers" or "not-poor-drivers" by the diagnostic test (neuropsychological test i.e. UFOV3). Only one "not-poor-driver" out of 48 was classified as a "poor-driver" by the test and three "poor-drivers" out of 8 as "not-poor-drivers". In total, out of the 56 drivers only 4 drivers were misclassified by the test

Table 6 The Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value and Correct Classification of the UFOV3 test

		Driver status (from Driving simulator)	
		Poor Driver (positive)	Not-poor Driver (Negative)
Test (Ufov3)	Test is positive (poor driver)	5	1
	Test is negative (Not-poor driver)	3	47
Total		8	48

Sensitivity = 5/(5 + 3) = 62.5%

Specificity = 47/(1 + 47) = 97.92%

Positive Predictive Value = 5/(5 + 1) = 83.33%

Negative Predictive Value = 47/(3 + 47) = 94%

Correct Classification = (5 + 47)/56 = 92.86%

(neuropsychological test i.e. UFOV3). Positive and negative predictive values of the test were high.

In order to determine the extent to which age alone can discriminate between "poor-drivers" and "not-poor-drivers", a logistic regression model using age as a predictor was developed (Eq. 3) and was subjected to diagnostic checks/goodness-of-fit test and confirmed through model validation.

$$\hat{\pi} = \frac{1}{1 + e^{-(-12.49311 + 0.1654765 \, age)}} \tag{3}$$

Where π is the probability of the outcome of interest (i.e. probability of being a "poor-driver"). An Odds Ratio (based on a unit increase in age) of 1.179955 and 5.23 (based on a 10 year increase in age) was obtained. Tables 4 and 5 show standard errors and confidence intervals for coefficient of age and odd ratio (based on unit increase in predictor) of the model. An area under ROC curve (Receiver Operating Characteristic curve) of 0.9062 was obtained which is considered excellent discrimination [21]. A cut-point of 0.56 provided the highest correct classification rate of drivers and resulted in the following figures: (i) Sensitivity = 50%, (ii) Specificity = 97.92%, (iii) positive predictive value = 80%, negative predictive value = 92.16%, correct classification = 91.07%. The cut-point value of predicted probability of 0.56 when adopted in Equation 3 corresponds to a cut-point value of 77 years (after rounding) of age. 91.07% of the drivers were correctly classified as either being "poor-drivers" or "not-poor-drivers" by the diagnostic test (i.e. age). Only one "not-poor-driver" out of 48 was classified as a "poor-driver" by the test and four "poor-drivers" out of 8 as "not-poor-drivers". In total, out of the 56 drivers only 5 drivers were misclassified by the test (i.e. age). Positive and negative predictive values of the test were high.

To explore if a cognitive measure based on a composite score of all nine neuropsychological tests (i.e. *Tscore_cognitive*) was a better discriminator than the UFOV3 test alone, a univariate logistic regression model based on *Tscore_cognitive* was developed. This model (Equation 4) was subjected to diagnostic checks/goodness-of-fit test and confirmed through model validation.

$$\hat{\pi} = \frac{1}{1 + e^{-(-12.91376 + 0.1067349 \, Tscore\_cognitive)}} \tag{4}$$

Where π is the probability of the outcome of interest (i.e. probability of being a "poor-driver"). An Odds Ratio (based on a unit increase in *Tscore_cognitive*) of 1.112639 and 2.91 (based on a 10 unit increase in *Tscore_cognitive*) was obtained. Tables 4 and 5 show standard errors and confidence intervals for coefficient of *Tscore_cognitive* and odd ratio (based on unit increase in predictor) of the model. An area under ROC curve (Receiver Operating Characteristic curve) of 0.9010 was obtained which is considered excellent discrimination [21]. A cut-point of 0.49 provided the highest correct

classification rate of drivers and resulted in the following figures: (i) Sensitivity = 50%, (ii) Specificity = 95.83%, (iii) positive predictive value = 66.67%, negative predictive value = 92%, correct classification = 89.28%.

That is, 89.28% of the drivers were correctly classified as either being "poor-drivers" or "not-poor-drivers" by the diagnostic test (i.e. Tscore_cognitive). Only two "not-poor-driver" out of 48 were classified as a "poor-driver" by the test and four "poor-drivers" out of 8 as "not-poor-drivers". In total, out of the 56 drivers 6 drivers were misclassified by the test (i.e. Tscore_cognitive). Positive predictive value was relatively not high but negative predictive value of the test was high.

4 Discussion on results

The UFOV3 test had the highest discriminating ability in separating poor-drivers from not-poor-drivers. The next best discriminating ability in decreasing order of strength is that of dichotic, trail, rey-copy and paper. This highlights the relevance of visuospatial skills and attention in gauging risky driving behavior, as the UFOV test primarily evaluates visual processing speed and divided and spatial attention. Although the UFOV3 test resulted in high Specificity but it furnished a relatively lower Sensitivity. The lower Sensitivity probably resulted from the fact that our sample consisted of active drivers from the general driving population and not from a clinical population; this sample apparently had good mental and physical constitution (and therefore, the differences between the younger group and the older group were much subtler). Had the older group come from a clinical population, the effects would have appeared stronger with consequent higher Sensitivity for the UFOV test. Some misclassification by using the test i.e. UFOV3 is bound to happen no matter how we ascertain the cut-point because of the overlap of the distributions of the "poor-drivers" and the "not-poor-drivers" i.e. some "poor drivers" will tend to have UFOV3 scores lower than "not-poor-drivers" while some "not-poor-drivers" will have UFOV3 scores higher than "poor drivers" (note: higher UFOV3 scores translate to lower cognitive status, see Table 3). This compromise will happen because we are using a simpler test (i.e. the UFOV3 test) as a proxy for a more elaborate, time consuming, expensive and accurate test (i.e. the driving simulator test) for ascertaining "poor-drivers" with the understanding that some misclassification will result. In practice, the test is most likely to be applied in a clinical setting, thereby giving much higher Sensitivities.

Similarly, age test also resulted in relatively lower Sensitivity. However, age cannot solely be made as a criterion for the discrimination of drivers as its effects are confounded by neurological diseases. Also, certain medical conditions/neurological diseases that have a tendency to bring about cognitive impairment to the extent that safe operation of motor vehicles is not possible and increasingly, such medical conditions have started to afflict people at relatively early ages. Even a small but significant number of younger people suffer from dementias that are likely to drive a motor car.

The composite neuropsychological measure (Tscore_cognitive) is not a better discriminator than the UFOV3 test alone in separating "poor drivers" from "not-poor-drivers". Using this composite neuropsychological measure as a discriminator, four "poor-drivers" and two "not-poor-driver'" were misclassified. Even age as a discriminator is slightly better in performance than this composite neuropsychological measure. It shows that the UFOV3 test alone is tapping relevant cognitive constructs (with regard to driver discrimination) compared to the "test all" cognitive skills approach that is being exercised through the composite neuropsychological measure. Also, this is testament to the much needed parsimony/economy in neuropsychological testing for driving and implies that preliminary driving-status can be determined without extensive investment in time (UFOV takes 15 min to administer).

The neuropsychological tests UFOV3, dichotic and rey-recall as a group emerged as the best predictors of a general driving skills index in this research (i.e. Eq. 1). This index is a measure of the general driving skill of a driver with essentially the same emphasis being placed on each driving performance parameter and therefore cannot be used to assess risky driving behavior due to neuropsychological impairment. However, it is a useful general index that can be used to gauge driving proficiency.It is reiterated that this research involved evaluation of driving performance as gauged by a driving simulator alone and no recourse was made to on-road driving test. Significant correlations have been found between the on road driving index and the driving simulator [8]. Participants in both the younger group and the older group had no previous experience in driving a driving simulator. Compared to the sample of 56 drivers in this work, a larger sample would have further enriched/improved the models.

The clock drawing test (clock) and the trail making test (trail) often used to clinically assess dementure did not emerge as significant predictors of driving ability in this research. Both tests are quick and easy to administer (are paper-and-pencil tests) and are recommended by the American Medical Association (AMA) for screening unsafe drivers. However, studies [36, 42] have shown the clock drawing test not to be a good screening instrument for detecting the very earliest signs of dementia. One potential draw-back of especially using the trail test for such predictive purposes is that a candidate could get hold of a standard testing sheet and through practice acquaint himself with the spatial configuration of

letters and alphabets which would allow him to get a higher but biased score.

A larger sample size and cognitive status evaluation of drivers (through clinical/hospital evaluation) was precluded because of limited financial resources; although recourse to such a strategy would have immensely enriched this research and would have enabled the framing of conclusions in a much favorable manner. Further research should encompass a much larger sample of drivers along with their cognitive status (through qualified clinicians) evaluation before recommendations can be framed for driver screening.

5 Conclusions

Acknowledging the limitations (small sample size and lack of cognitive status), our research highlights are as under:

1) For a UFOV3 score greater than 220, the driver may be evaluated further by a driving specialist to ascertain questionable driving behavior.
2) Drivers above the age of 77 were more susceptible to exhibiting unusual driving behavior.
3) In place of currently used tests, the UFOV test may show more promise in driver evaluation.
4) On average cluster analysis/driving performance and neuropsychological tests show comparatively low scores for poor drivers (group 2).

Further research using a larger sample of new drivers (both young and old) will confirm results before any firm conclusions can be drawn on practical implications with regard to driver screening.

Abbreviations AIC, Akaike Information Criterion; AMA, American Medical Association; BIC, Bayes Information Criterion; DAT, Dementia of the Alzheimer Type; DPI, Driving Performance Index; DA, Divided Attention; DVLA, Driver and Vehicle Licensing Agency; ROC, Receiver Operating Characteristic; UFOV, Useful Field of View; UK, United Kingdom

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Allen RW, Marcotte TD, Rosenthal TJ, Aponso BL (2005) Driver Assessment with Measures of Continuous Control Behaviour. Proc. 3rd Intl. symp. on Human Factors in Driver Assessment, Training, and Vehicle Design. Rockport, Maine
2. Anderson SW, Rizzo M, Shi Q, Uc EY, Dawson JD (2005) Cognitive Abilities Related to Driving Performance in a Simulator and Crashing on the Road. Driving Assessment 2005: 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Rockport, Maine
3. Ball KK, Vance DE, Edwards JD, Wadley VG (2004) Aging And The Brain. In: Rizzo M, Eslinger PJ (eds) Principles and Practice of Behavioral Neurology and Neuropsychology. Saunders-An imprint of Elsevier, Philadelphia, pp 795–809
4. Box E, Gandolfi J, Mitchell K (2010) Maintaining safe mobility for the aging population–The role of the private car. London: RAC Foundation. Retrieved from: http://www.racfoundation.org/assets/rac_foundation/content/downloadables/maintaining%20safe%20mobility%20-%20rac%20foundation%20-%20140410%20-%20report.pdf
5. Brown LB, Ott BR (2004) Driving and Dementia: A Review of the Literature. J Geriatr Psychiatry Neurol 17:232–240
6. Carr DB, O'Neill D (2015) Mobility and safety issues in drivers with dementia. Int Psychogeriatr 27(10):1613–1622
7. Carter K, Monaghan S, O'Brien J, Teodorczuk A, Mosimann U, Taylor JP (2015) Driving and dementia: a clinical decision pathway. Int J Geriatr Psychiatry 30(2):111–222
8. Casutt G, Martin M, Keller M, Jancke L (2014) The Relation Between Performance in On-Road Driving, Cognitive Screening and Driving Simulator in Older Healthy Drivers. Transp Res F 22(2014):232–244
9. Cheung I, McCartt AT (2011) Declines in fatal crashes of older drivers: Changes in crash risk and survivability. Accid Anal Prev 43(2011):666–674
10. Christie N, Savill T, Buttress S, Newby G, Tyerman A (2001) Assessing Fitness to Drive After Head Injury: A Survey of Clinical Psychologists. Neuropsychol Rehabil 11(1):45–55
11. Dellinger AM, Kresnow M, White DD, Sehgal M (2004) Risk to self versus risk to others: How do older drivers compare to others on the road? Am J Prev Med 26(3):217–221
12. Driver & Vehicle licensing Agency (DVLA) (2017) Assessing fitness to drive– a guide for medical professionals; October 2017. www.gov.uk/dvla/fitnesstodrive
13. Evans L (2000) Risks older drivers face themselves and threats they pose to other road users. Int J Epidemiol 29(2):315–322
14. Fraley C, Raftery AE (2006) MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering. Technical Report No. 504. Department of Statistics, University of Washington. Revised (minor) November 2007
15. Groeger JA (2016) Understanding Driving: Applying Cognitive Psychology to a Complex Everyday Task. Psychology Press. Taylor & Francis Group, Sussex
16. Hakamies-Blomqvist L (1993) Fatal Accidents of Older Drivers. Accid Anal Prev 25(1):19–27
17. Hakamies-Blomqvist L, Raitanen T, O'Neill D (2002) Driver ageing does not cause higher accident rates per km. Transp Res F 5(2002):271–274
18. Hakamies-Blomqvist L, Sirén A, Davidse R (2004) Older drivers – a review. VTI rapport 497A. VTI, Linköping
19. Harris PB, Keady J (2004) Living With Early Onset Dementia: Exploring the Experience and Developing Evidence-based Guidelines for Practice. Alzheim Care Today 5(2):111–122
20. Harris PB, Keady J (2009) Selfhood in younger onset dementia: Transitions and testimonies. Aging Ment Health 13(3):437–444

21. Hosmer DW, Lemeshow S, Sturdivant RX (2013) Applied Logistic Regression, 3rd edn. Wiley, New York

22. Keall MD, Frith WJ (2004) Older Driver Crash Rates in Relation to Type and Quantity of Travel. Traffic Inj Prev 5(1):26–36

23. Khan MT (2009) The Effects of Ageing on Driving Related Performance. School of Civil Engineering and the Environment. University of Southampton, UK. Doctoral Thesis. Available at http://eprints.soton.ac.uk/73700/

24. Kline P (2016) A Handbook of Test Construction: Introduction to Psychometric Design. Routledge, London and New York

25. Kutner MH, Nachtsheim CJ, Neter J, Li W (2005) Applied Linear Statistical Models. 5/e. McGraw-Hill, New York

26. Langford J, Koppel S, McCarthy D, Srinivasan S (2008) In Defence of 'Low-mileage' bias. Accid Anal Prev 40(6):1996–1999

27. Leeuwen PMV, Subils CGI, Jimenez AR, Happee R, De Winter JCF (2015) Effects of visual fidelity on curve negotiation, gaze behaviour and simulator discomfort. Ergonomics 58(8):1347–1364

28. Martin AJ, Marottoli R, O'Neill D (2013) Driving assessment for maintaining mobility and safety in drivers with dementia. Cochrane Database of Syst Rev (8): CD006222. https://doi.org/10.1002/14651858.CD006222

29. McKenna P (1998) Fitness to Drive: A neuropsychological perspective. J Ment Health 7(1):9–18

30. Meuleners LB, Harding A, Lee AH, Legge M (2006) Fragility and crash over-representation among older drivers in Western Australia. Accid Anal Prev 38:1006–1010

31. Michels TC, Tiu AY, Graver CJ (2010) Neuropsychological Evaluation in Primary Care. Am Fam Physician 82(5):495–502

32. Mourant RR, Thattacherry TR (2000) Simulator Sickness in a Virtual Environments Driving Simulator. Proceedings of the IEA 2000/HFES 2000 Congress

33. Parasuraman R, Nestor PG (1993) Attention and Driving: Assessment in Elderly Individuals with Dementia. Clin Geriatr Med 9(2):377–387

34. Patomella A, Kottorp A (2005) An Evaluation of Driving Ability in a simulator: A good Predictor of Driving Ability after Stroke? Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Rockport, Maine

35. Pavlou D, Beratis D, Fragkiadaki S, Kontaxopoulou D, Yannis G, Economou A, Papageorgiou S (2016) Which are The Critical Parameters Assessing the Driving Performance of Drivers With Cerebral Diseases? A Literature Review. World Conference on Transport Research - WCTR 2016 Shanghai

36. Powlishta KK, Dras VDD, Stanford A, Carr DB, Tsering C (2002) The clock Drawing Test is a Poor Screen for Very Mild Dementia. Neurology 59:898–903

37. Reger MA, Welsh RJ, Watson GS, Cholerton B, Baker LD, Craft S (2004) The Relationship Between Neuropsychological Functioning and Driving Ability in Dementia: A Meta-Analysis. Neuropsychology 18(1):85–93

38. Siren A, Haustein S (2015) Driving licences and medical screening in old age: Review of literature and European licensing policies. J Transp Health 2(1):68–78

39. Staplin L, Lococo KH, Stewart J, Lawrence E (1999). Safe Mobility For Older People. Report DOT HS 808 853. National Highway Traffic Safety Administration. U.S. Department of Transportation. At http://www.nhtsa.dot.gov/people/injury/olddrive/safe/tech-doc.htm

40. Stinchcombe A, Paquet S, Yamin S, Gagnon S (2016) Assessment of Drivers with Alzheimer's Disease in High Demand Driving Situations: Coping with Intersections in a Driving Simulator. Geriatrics 1(21). https://doi.org/10.3390/geriatrics1030021

41. STISIM (2009) Systems Technology Incorporated, 13766 South Hawthorne Blvd. Hawthorne, CA 90250–7083. The version of the software was Build 2.08.01 copyright © 1985–2009

42. Storey JE, Rowland JT, Basic D, Conforti DA (2001) A comparison of Five Clock Scoring Methods using ROC (receiver operating characteristics) Curve Analysis. Int J Geriatr Psychiatry 16:394–399

43. Thompson KR, Johnson AM, Emerson JL, Dawson JD, Boer ER, Rizzo M (2012) Distracted driving in elderly and middle-aged drivers. Accid Anal Prev 45(2012):711–717

44. Whelan M, Langford J, Oxley J, Koppel S, Charlton J (2006) The Elderly and Mobility: A Review of the Literature. Report No. 255, Monash University, Accident Research Center. Retrieved from: http://www.monash.edu.au/miri/research/reports/muarc255.pdf