The relation between performance in on-road driving, cognitive screening and driving simulator in older healthy drivers

Gianclaudio Casutt a,b,c,d,* , Mike Martin b,d,e , Martin Keller c , Lutz Jäncke c,d,e,f

a Department of Human Sciences, University in the Principality of Liechtenstein, Dorfstrasse 24, 9495 Triesen, Liechtenstein
b Department of Psychology, Division of Gerontopsychology, University of Zurich, Switzerland
c Department of Psychology, Division of Neuropsychology, University of Zurich, Switzerland
d International Normal Aging and Plasticity Research Centre (INAPIC), University of Zurich, Switzerland
e University Research Priority Program “Dynamics of Healthy Aging”, University of Zurich, Switzerland
f King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

As people age, physical and psychological functions deteriorate which is associated with higher crash risk. In different countries traffic authorities developed screening systems to identify unsafe older drivers. However, until today these screening systems (medical check, road test, cognitive screening) failed. In this context, driving simulators could be helpful in measuring driving performance under controllable and replicable experimental conditions in near-crash situations. However, little is known about the relation between older drivers’ performance in an on-road test, cognitive screening and driving simulator performance. In this paper we examined the relation between these three driving related measures in a sample of older participants (mean age = 72.12 yrs.). The results show that performance in an on-road test and the performance in the cognitive tests explained 50% of the variance in the driving simulator performance (r = .71). The current analysis indicates that driving simulator data represents on-road driving behaviour and cognitive performance of older drivers. This research presents a substantial potential for driving simulators, for example identifying or retraining unsafe older drivers.

1. Introduction

1.1. Older driver and driving safety

Older adults are the fastest growing driver segment in industrial countries in the next ten years. For 2020 it is estimated that between 33 and 38 million, or 25% of older drivers will be part of the driving population in the US. Highest fatality risk is found in driver aged 80 and older and their involvement is estimated to increase by 155% by 2030 (Lyman, Ferguson, Braver, & Williams, 2002). Nevertheless an increase of fatality risk in older drivers has not been observed. Cheung & McCartt, 2011 showed that fatality risk in drivers aged 75 and older per 100,000 licensed drivers was higher than for drivers aged 35–54 over the period from 1997 to 2008. Interestingly, in the same period the reduction of fatal crash risk in the group aged 80 and older was greater than for the group aged 35–54. Environmental improvements in an emergency, vehicle and traffic safety...
Comparison of age dependant crash risk is difficult because the vehicle, exposure, situation and, the most complex, the drivers’ behaviour differs in age groups and influence the crash risk (Classen et al., 2007). But physical and psychological functions deteriorate with age and are associated with an increased crash risk in older drivers (Anstey, Wood, Lord, & Walker, 2005; Emerson et al., 2012). Therefore screening systems were developed to identify unsafe older drivers. A recent published study highlighted that so far the screening systems which used medical check-ups, on-road tests and cognitive screenings fail to enhance traffic safety. In the observed time period there was no reduction in fatality of older drivers but an increased involvement in unprotected older road users (Siren & Meng, 2012). Thus, coping with age-dependent decline in driving performance continues to be a substantial challenge for the community (Dobbs, 2008). Therefore it will be of outmost importance to develop techniques and strategies to identify unsafe drivers while keeping traffic safety (Dawson, Uc, Anderson, Johnson, & Rizzo, 2010). As driving cessation reduces well-being in old age (Fonda, Wallace, & Herzog, 2001; Marrottoli et al., 2000) there is also a need to develop strategies to maintain driving performance in older drivers (Roenker, Cissell, Ball, Wadley, & Edwards, 2003).

1.2. Driving performance and cognitive performance

Typical age-related changes in driving performance are caused by declines in sensory (vision and auditory), cognitive and physical abilities (Baldock, Thompson, & Mathias, 2008; Ball, Owsley, Sloane, Roenker, & Bruni, 1993; Edwards, Bart, O’Connor, & Cissell, 2009; Owsley & Ball, 1993; Owsley et al., 1998). So far a vast number of computer-based cognitive tasks have been designed to assess driving performance. As mentioned before, the usability of these tests in identifying older unsafe drivers remains uncertain. They disagree from each other and there is either a lack of cut-off scores, or sensitivity and specificity to identify unsafe drivers are unknown (Langford & Koppel, 2006; Mathias & Lucas, 2009; Poschadel et al., 2012). Furthermore there is also a lack of reliable and valid standardized data from on-road assessments reflecting driving performance especially in old age (Selander, Lee, Johansson, & Falkmer, 2011). Therefore some researchers propose a multi-tiered assessment including on-road tests and cognitive tests as well as driving simulator settings (Langford et al., 2008; Lees, Cosman, Lee, Fricke, & Rizzo, 2010).

Risser et al. (2008) analysed the relation between an on-road test and the performance on a computer-based cognitive test battery in a sample of young and middle-aged drivers. The cognitive measures are mathematically combined to one composite measure, which was, based on a trained artificial neural network (logistic regression approaches), statistically related with the on-road test. They identified six cognitive tests (selective and divided attention, field of view, perceptual speed, physical motor speed and fluid intelligence) explaining 50% of the variance in the on-road test (Risser et al., 2008). However, a further study found only weak correlations between cognitive tests and two different on-road tests. Only the UFOV was significantly correlated with both on-road tests with a maximal variance of 21%. The authors of this study argue that the lack of validity and reliability of the on-road test driving measures are pivotal for the weak correlation (Selander et al., 2011).

Mathias and Lucas (2009) focused in their meta-analysis on the relation between age dependent decline in cognition and driving performance. Performances in different cognitive tests were related to driving performance in on-road driving, simulator driving and driving offences. Only test scores of UFOV test met their study criteria for predicting older drivers driving performance. The predicting variance of the UFOV was 43% for the on-road driving performance, 36% for the simulator performance and 20% for the driving offences. However, they did not assess the direct relation between the on-road test and driving simulator performance.

1.3. Potential of driving simulators

A study using simulated driving scenarios employed hundred-twenty-nine older participants (60–88 years old) in a relatively simple driving simulator and validated driving simulator performance by an on-road test conducted on a predefined route (Lee, Cameron, & Lee, 2003). The on-road test performance was rated by driving experts while the driving simulator performance was rated half by an experimental assistant and half by automatically obtained measures recorded by the simulator software (driving speed, use of indicator, decision and judgement, confidence on high speed and attention task). The on-road test predicted 67% of the variance in the simulator performance. Both measures were negatively associated with chronological age ($-0.25 < r > -0.66$). Another study examined differences in cognitive performance and driving simulator performance in 20 younger (25–42 years) and 20 older drivers (65–83 years). 12 older drivers were excluded because of Simulator Adaptation Syndrome (SAS). Comparisons of challenging traffic events (car overtaking, pedestrian crossing, incursion of a parked car) during simulator driving showed higher crash risk for the older driver group. In addition, participants who crashed scored lower on the cognitive tests (reaction time, choice reaction time, and UFOV). The comparison of older non-crashers with older crasher showed a lower performance on the divided attention and selective attention UFOV subtests for the older crashers. These findings suggest a problematic role of older drivers’ cognitive capacity in demanding traffic situations (Bélanger, Cagnon, & Yamin, 2010).

Driving simulators seem to be a useful alternative to on-road tests since they offer the potential to design standardized driving scenarios which are the basis to obtain good measures. A study with older healthy participant and participants with
macular degeneration showed no differences in cognitive performance but different driving performance on a simulator and an on-road test (Szyk et al., 1995), which underlines the validity of driving simulator instruments. Wang et al. (2010) concluded in their study that driving simulators are a safe method for assessing driving performance providing valid estimates of on-road vision distraction as in-vehicle interface interactions.

There is no study published comparing the driving performance in an on-road test, cognitive tests and driving performance in a simulator test, especially for healthy older participants. This is somewhat surprising since these three measurement domains have all been used separately to test driving performance in older drivers (Ball, Edwards, & Ross, 2007; Burdick et al., 2003; Bélanger et al., 2010; Cassavaugh & Kramer, 2009; Edwards, Myers, et al., 2009; Kay, Bundy, & Clemson, 2009; Lee et al., 2003; Vance et al., 2006). However, driving simulator tests used mostly have employed simple traffic situations, which do not reflect the typical traffic situations in larger European cities. In addition, the problematic behaviour of older drivers seems to be located in demanding and complex traffic events (Vance et al., 2006) and was showed to end more often with a crash for those with reduced cognitive abilities (Bélanger et al., 2010). In a review of the literature, Lees et al. (2010) offered an interesting perspective of the usability of a driving simulator: while standardized neuropsychological tests are capable of measuring cognitive performance with maximum experimental control but minimal real life applicability, on-road tests are associated with minimal experimental control but maximum real life association. In conclusion a driving simulator test is in between of these two turning points and could offer a compromise of these two measurements approaches (Lees et al., 2010).

Other authors have noted an iceberg analogy describing the relation between older drivers’ driving behaviour and safety errors, meaning that the registered driving errors represent only reported accidents as the tip-of-the-iceberg, whereas the more common dangerous near-crash events (individual failed hazard perception (Horswill et al., 2008)) are below-the-waterline and therefore not registered. For maintenance of on-road safety it is thus necessary to identify unsafe older drivers with evidence-based on- and/or off-road assessments (Aksan et al., 2012). Reduced hazard perception is associated with reduced age related deficits in cognitive functioning (Borowsky, Shinar, & Oron-Gilad, 2010; Horswill et al., 2008). According to existing literature a driving simulator could be a new approach to identify unsafe older drivers because of its possibility to simulate and provoke near-crash events. The simulator shows driving reality under controlled and harmless conditions. It should be emphasised once again that in different countries on-road and cognitive screenings are implemented with the aim to identify unsafe older drivers (also in Switzerland, where the study was conducted). As mentioned, Siren and Meng (2012) pointed out that these screening policies failed. However it is possible that the low number of unsafe older drivers is able to compensate their deficits during the screening policies because of the lack of critical events.

1.4. Aim of this paper

There is need to study driving performance in more complicated traffic situation especially for older drivers since complexity of cognitive control is an important variable in driving performance as well as in aging (Aksan et al., 2012). Furthermore the driving simulator scenarios should include hazardous and normal traffic situations with high reference to reality. Different to the study by Lee and colleagues (Lee et al., 2003) we used a driving simulator offering more complex and hazardous traffic situations. Also different from the study of Lee and colleagues we focus on rural roads and urban streets to summarise a daily road trip with reference to reality. So we implemented the normal simulator situations from the study of Lee et al. (2003) and the hazardous simulator situations from Bélanger et al. (2010). However, it is of outmost importance to examine whether the overall performance of the driving simulator measures similar, shared or different aspects of the two other measurements (on-road test and cognitive screening) in older drivers. Since driving simulator performance is of outmost importance for this paper, we will use driving simulator performance as dependent variable in this paper.

2. Materials and methods

2.1. Participants

The participants were recruited by a newspaper article and a newspaper advertisement. A total of 244 participants indicated interest to participate in this study. All of them received detailed study information and a checklist screening the active driving status (kilometres/year, age receiving driving licence, medication influencing driving, visual problems). Participants were included if they did not suffer from neurological and psychiatric diseases, which was evaluated using a short medical checklist during recruiting. 91 participants fulfilled all inclusion criteria and agreed to participate in the study. All participants had an original valid driver licence. In Switzerland drivers older than 70 years must participate every two years in screening police (medical and cognitive screening) for renewal of their drivers’ licence. Based on the study design 26 participants did not perform the driving simulator test and are not included in the final analysis. They enrolled in a different study as control participants. 13 participants were excluded from the analysis due to reported simulator sickness (SS), 3 participants stopped their participation because of personal problems. A sample of 49 participants was thus included in the final analysis (Table 1).
2.2. Ethical considerations

This project (Drive Wise) was approved by the Cantonal Ethic-Commission of Zurich, University Hospital of Zurich (KEK-ZH-NR: 2010-0090/0). The traffic and police departments gave official permission to conduct the on-road test assessment. A signal box was placed on top of the private car of each participant during the on-road test. According to the rules provided by the ethics committee all participants received written study information and were informed that participation would not impinge on their driving licence. All participants were explicitly instructed that they were able to terminate the study at any time without any negative consequences.

2.3. Study protocol and procedures

The Drive Wise project was conducted in the context of a larger research project focussing on brain plasticity and practicing driving performance in older participants. In this paper we will report the relation between the performances of on-road test, cognitive test and simulator test. The training data will be reported elsewhere.

The study participation for each participant was seven weeks. In the first and the seventh week participants were tested with an on-road test and a cognitive test battery. In the second, fourth and sixth week participants were tested once on a driving simulator. A commonly reported problem when using a driving simulator is that the participants report simulator sickness (SS) quite frequently. During the three simulator sessions SS decreased substantially in the examined participants because they gained experience with the simulator situation. Thus, for driving simulator performance we took mean measures of repeated driving simulator sessions in order to receive a stable driving simulator measure more or less uninfluenced by simulator sickness. 13 participants did not enroll in final analysis because of SS. There were no significant differences in demographic variables in participants with manifesting SS from participants who did not. There were significant gender differences. Women showed more SS than men.

Because of severe SS in the first driving simulator session in some participants we calculated only from the second and third driving simulator sessions the mean measures of each participant. Similar, we also calculated a mean for the on-road test and cognitive measures obtained in the two sessions before and after the driving simulator sessions. For on-road and driving simulator assessment a factor analysis (principal components analysis) examined the mean measures and reduced them to factors. For the cognitive assessment the software directly calculated one factor (Schuhfried, 2005).

2.4. Simulator sickness score

Global SS was measured by the mean of the three main symptoms: nausea (N), oculomotor (O) and disorientation (D) (Kennedy, Lane, Berbaum, & Lilienthal, 1993) in the three simulator driving sessions. Each symptom was scored from 1 to 5 (low SS = 1, severe SS = 3, strong SS = 5). Friedman’s ANOVA was used to compare the global SS scores across the driving simulator sessions. The global SS of participants significantly changed during the driving simulator session within the 5 weeks ($\chi^2 = 34.29, p < 0.001$). Post-hoc Bonferroni corrected Wilcoxon tests revealed a drop in subjectively experienced simulator sickness from the start of the driving simulator session (median = 2.33) to half time of the driving simulator session (median = 1.33, $z = -4.58, p < .001$). There is also a significant reduction of subjectively experienced SS from the start to the end of driving simulator session (median = 1.33, $z = -4.71, p < .001$). Global SS did not significantly change from half time of session (median = 1.33) to the end of simulator participation (median = 1.33, $z = -0.27, p = .79$).

2.5. Assessments and analysis

2.5.1. On-road test driving assessment

The on-road test drive took approximately 45 min on a prescribed route (24.9 km) on public roads including district and urban streets, suburb and rural roads and a motorway passage. The test track was chosen in collaboration with driving experts, who use this route on a regular basis for their on-road test exercises. A driving instructor (DI) sat beside the participant, supervised driving safety, observed different aspects of the driver’s behaviour, and rated the driving performance after finishing the driving session. The Zurich On-road test Assessment sheet (ZOA) is a modified version of the formal evaluation sheet used by the DI. The DI was instructed to evaluate cognitive aspects of driving behaviour but not the performance with respect to parking manoeuvres and car handling. Seven different aspects (Table 2) were included in the ZOA, which were

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Age (years)</th>
<th>Gender</th>
<th>Years of license</th>
<th>km/year</th>
<th>Gearbox</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (SD)</td>
<td>Male</td>
<td>Female</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>No. SS</td>
<td>49</td>
<td>72.12 (5.40)</td>
<td>36</td>
<td>13</td>
<td>50.33 (5.67)</td>
<td>11238.78 (6520.79)</td>
</tr>
<tr>
<td>SS</td>
<td>13</td>
<td>73.62 (4.33)</td>
<td>6</td>
<td>7</td>
<td>49.77 (6.15)</td>
<td>9538.46 (5796.77)</td>
</tr>
</tbody>
</table>

SS = Simulator Sickness, A = automatic, M = manual.
evaluated by the DI. All dimensions were rated on a scale including six to eight 5-point-items (poor = 1, sufficient = 3, excellent = 5). The DI was blinded with respect to the study design. The principal investigator or an assistant sat in the rear seat to control that the participants and the DI did not talk with each other about the study during driving.

The z-transformed variables from the ZOA were subjected to a factor analysis using varimax rotation (all prerequisites for performing the factor analysis were fulfilled: Kaiser–Meyer–Olkin measure = .71, Bartlett’s test of sphericity $X^2(21) = 258.38, p < .001$). The first factor of this analysis explained 62.27% of the variance. Factor scores were computed for this factor representing the driving performance. Based on these factor scores an index was calculated representing the overall driving performance (the on-road test index). A larger on-road test index indicates a better driving performance. The formula for calculating the on-road test index with the factor weights estimated from the factor scores is:

$$\text{On-road test index} = 0.190 \times \text{observation and interaction} + 0.200 \times \text{behaviour in turnaround} + 0.167 \times \text{gaze behaviour} + 0.115 \times \text{use of direction indicator} + 0.210 \times \text{district dependant behaviour} + 0.183 \times \text{use of different speed limit} + 0.191 \times \text{lane behaviour}$$

2.5.2. Cognitive assessment

Cognitive abilities were tested with a standardised computer-based test battery of the Expert System Traffic XPSV (Schuhfried, 2005). This test battery is known in the Austrian and German traffic psychological field and is often used for traffic scientific questions depending on reduced cognitive abilities (Sommer, Herle, et al., 2008; Sommer et al., 2010). As described in chapter 1.2 this test battery explained 50% of the variance in the data of a sample in a on-road test (for more detail see also Risser et al. (2008)). The following cognitive subtests (Table 3) were included:

The Reaction Test (RT) is a simple choice reaction time task. Three different stimuli (yellow or red circle and an acoustic signal) are used in this task. Participants were required to react to the simultaneous presentation of the yellow circle (presented on the screen) and the pitched sound by pushing a corresponding target button with the right index finger. In all other conditions the participants had to withhold reactions and keep the finger placed on the corresponding start button. Decision speed (DS) is measured by the latency from stimulus onset until lifting off the start button while the physical motor speed (MS) is defined as the movement time from the start button to the target button. The reliability coefficients (DS, MS) amount to .94 and .98, respectively.

The Cognitrone Test (COG) measures selective attention. Six blocks of different complex stimuli (total 60 trials) are presented on the screen. The upper part of the screen four constantly presented reference stimuli (10 trials per block) are

<table>
<thead>
<tr>
<th>Test</th>
<th>Cognitive paradigm</th>
<th>Output variables</th>
<th>Reliability</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction Test (RT)</td>
<td>Simple choice reaction</td>
<td>Reaction time decision and motor speed (DS, MS)</td>
<td>DS = .94</td>
<td>Schuhfried and Prieler (1997)</td>
</tr>
<tr>
<td>Cognitrone (COG)</td>
<td>Selective attention</td>
<td>Correct reactions, required processing time</td>
<td>MS = .98 ,95</td>
<td>Wagner and Karner (2001)</td>
</tr>
<tr>
<td>Determination Test (DT)</td>
<td>Complex choice reaction</td>
<td>Angular dimension, number of detection</td>
<td>.98 FV = .96 DA = .98</td>
<td>Schuhfried (1998)</td>
</tr>
<tr>
<td>Peripheral Perception (PP)</td>
<td>Field of view (FV), divided attention (DA)</td>
<td>Correct response</td>
<td>.80</td>
<td>Schuhfried, Prieler, and Bauer (2002)</td>
</tr>
<tr>
<td>Tachistoscopic Traffic Perception Test (ATAVT)</td>
<td>Perceptual speed</td>
<td>Correct answer</td>
<td>.80</td>
<td>Sommer, Herle, Häusler, and Arendasy (2008)</td>
</tr>
<tr>
<td>Matrices Test (AMT)</td>
<td>Fluid intelligence</td>
<td>Correct answer</td>
<td>.70</td>
<td>Hornke, Etzel, and Rettig (2003)</td>
</tr>
</tbody>
</table>
shown. Below these reference stimuli a test stimulus is presented which alternates for each trial. Participants have to decide if the test stimulus is identical with one of the four reference stimuli by pressing one of two corresponding buttons. This test has no time limitation. Mean reaction time of correct incongruent answers (CIAn; comparison new stimulus with all four reference stimuli) serves as measurement of selective attention. The reliability coefficient (CIAn) amounts to .95.

The Determination Test (DT) is a complex choice reaction time task. Stimuli are five circles of different colours (white, red, green, yellow, blue), two chequered rectangles and a high and low pitched sound. The circles are presented on the upper part of the screen at undefined positions; the rectangles are presented on the left and right lower quadrants of the screen. All stimuli (visual and acoustic) are presented separately and in random order. Participants are required to push the corresponding buttons before presentation of the next stimulus. For the five circles there are representative coloured buttons, for the left and right presented chequered rectangle participants have to push a corresponding foot pedal (e.g.: left foot when rectangle in the left corner). For the high sound participants have to press a grey button, for the low sound a black button. The task duration is four minutes. The test is administered as a computerized adaptive test (CAT). The 1PL Rasch model person parameter estimate represents the main variable of this test. It is calculated taking into account the difficulty of the items administered and whether each item has been solved by the participant or not. In this task the CAT is conceptualized not as classic CAT. The stimuli presentation is presented a little faster than the optimal speed for participants. The number of correct responses (CR) was the main variable and represents sensory stress. The reliability coefficient (CR) amounts to .98.

The Peripheral Perception Test (PP) utilizes a field of view and divided attention paradigm. For the central visual field a horizontal moving red ball (reference value) on the screen has to be tracked with a moving cross (actual value). In the left and right peripheral visual field diode panels are placed inducing the perception of moving lights. From time to time a vertical line of diodes is presented in the periphery. Participants have to track the actual value with a control device by simultaneously observing the appearance of a vertical line in the periphery and pressing a foot pedal. The distance to the screen is measured by infrared camera (at least 60 cm). Divided attention (DA) is measured as the absolute deviation between actual and reference value. Visual field (VF) is measured as the widest field angle at which the vertical lines of diodes are detected with respect to the distance of the screen. The reliability coefficients (VF, DA) amount to 0.96 and 0.98.

The Adaptive Tachistoscopic Traffic Perception Test (ATAVT) is an object perception test. Traffic specific photographs of different complexities are presented for a short time (700–1300 ms). Participants have to decide if (1) vehicles, (2) bicycles, (3) pedestrians, (4) road signs, or (5) traffic lights are contained. The test is administered as a CAT. The number of correctly identified objects weighted by complexity of the figure is the dependent measure for perceptual speed (PS). The reliability coefficient amounts to 0.80.

The Adaptive Matrices Test (AMT) is a fluid intelligence (FI) test. The stimuli are comparable to classical matrices, but based on explicit construction rules. Participants have to identify the figural pattern among eight answers alternatives which conclude the presented matrices’ item. The test is administered as a CAT. The reliability coefficient (FI) amounts to .70.

In previous research the main variables of the above described cognitive test measures have been analysed by multivariate classification algorithms (artificial neural networks) that assign participants to predefined categories on the basis of one neural network (NN) variable (Risser et al., 2008). From these empirical results a global score for NN was evaluated, based on the Austrian and German practical on-road tests. Global scores of 4 or 5 indicate an insufficient driving behaviour so that the participant would fail the on-road test. Global scores of 1, 2, and 3, which correspond to very good, good, and satisfactory driving behaviour, indicate that the participant would pass the on-road test. (Risser et al., 2008; Sommer, Herle et al., 2008; Sommer et al., 2010). The NN validly estimated the composite score as demonstrated by a good jack-knife validity coefficient of R = .77. The composite score is denoted as cognitive index throughout this paper and is used as general indicator of cognitive function.

2.5.3. Driving simulator assessment

A commercial driving simulator was used for this study (Dr. Foerst; Trainer F12PT-1L40; 22” Samsung LCD-screen in a distance of approximately 100 cm to the steering wheel; software version 12, see also Jäncke and Klimmt (2011)). Participants drove in a single seated driver cabin of a Ford Focus® equipped with a steering wheel, a starter lock, a tachometer, signalers for light and blinker, wiper control switch, clutch, breaking and throttle pedals as well as gearshift. The software controlling the simulator and the traffic scenarios runs on Windows 7 operating system. The data collection was done with the maximum time resolution, which is equal to the frame rate of the screen.

The participants executed twice four different traffic scenarios (three rural and one urban) with oncoming traffic and curves. These drive through these scenarios lasted in total approximately 15 min. Participants were instructed to drive with adequate speed and follow the instructions of the “simulated trainer”. The “simulated trainer” was a computer-based program: a male voice giving information about the direction of travel. In all scenarios there was oncoming traffic. In three scenarios participants had to stop as fast as possible when an obstacle appeared. The rural scenarios were designed to simulate typical drives on rural roads, where the subjects have to keep an adequate speed (70 km/h), keep enough space to the cars driving ahead, and follow the direction of the road as precisely as possible. The driving distance varied between 0.8 and 2.8 km. In one of these scenarios (driving distance = 1 km) with few curves and some buildings a “STOP” signal appeared in the middle of the screen indicating to the participants to brake sharply. In a further scenario with more curves and trees (distance = 2.8 km) a deer crossed the road unexpectedly for three times forcing the driver to stop as fast and precisely as possible. In the third scenario with few curves, buildings and a parking area (distance = 0.8 km) subjects were instructed to follow the traffic rules as precisely as possible, and to stop for unexpected events (a child crosses the street, a parked
car opens the door or drove suddenly close to the driver’s front onto the road). A fourth scenario was a bit more complex (distance = 1.5 km). In this scenario different unexpected situations were randomly simulated (child with a ball jumps on the road, a pedestrian or cyclist hidden behind another object crosses the street, a parked car opens a door). As in the scenarios mentioned above the participants are instructed to stop the car as fast as possible in the event one of these obstacles occur and to avoid a clash. Reaction times were measured from the time of appearance of the particular event to the time of operating the brake. The computer software automatically elicited the insertions of the hazardous events. Reaction times longer than 1.5 s inevitably resulted in clashes.

The driving simulator software automatically recorded 6 variables (see Table 4) for each scenario. The 6 output variables obtained during the three rural and the urban street scenarios were significantly different from each other (p < .05; except driving errors p = .07). Participant dependent correlations of all variables in the rural and urban street scenarios were significantly and positively correlated (r = .24, p < .05). So the differences did not result from a difference in participants’ driving performance, but from differences in the scenarios (rural vs. urban street).

For this reason the rural variables and urban street variable were used as independent variables for the following PCA.

The z-transformed 12 driving simulator variables (6 variables for each rural and urban) were subjected to a factor analysis. One variable (lane variability in urban scenario) was excluded from the factor analysis, since the KMO statistic revealed that this variable was not sufficient for the factor analysis. There might have been an output variable problem in lane variability in location of intersections. For the remaining 11 variables a sufficient KMO statistic (.61) revealed that the data sample is adequate for performing a factor analysis. Bartlett’s test of sphericity (X^2 (55) = 195.67, p < .001) indicated that correlations between items were sufficiently large for PCA. The first factor explained 32.03% of the variance. From this factor, factor scores were computed representing the driving simulator performance. The formula for calculating the driving simulator index is given below:

- **Driving simulator index**: Rural scenario (0.031 top speed + 0.132 mean speed − 0.005 lane accuracy + 0.129 lane variability + 0.479 reaction time + 0.324 driving errors) + urban street scenario (−0.096 top speed + 0.111 mean speed − 0.216 lane accuracy + 0.260 reaction time + 0.300 driving errors)

### 2.6. Statistical analysis

All statistical analysis were performed using SPSS® (version 18) on a PC equipped with Windows 7 operating system. As a rule of thumb, the sample size for regression models with high level of power (Cohen’s $d = .8$ which refers 14% of Variance) and two predictor variables need a minimum of 40 participants (page 223 in Field, 2009). Before statistical testing the variables were tested for normal distribution. If not otherwise noted the data fulfilled all prerequisites for conducting the statistical analyses. Cronbach’s $\alpha$ coefficients were computed for the on-road, cognitive and simulator indices revealing moderate to strong coefficients (on-road = 0.89, cognitive = 0.69 and driving simulator = 0.54).

On-road index, cognitive index and driving simulator index were subjected to a hierarchical regression analysis with the on-road and cognitive-indices as predictors and the driving simulator index as dependent variable. The regression model was based on the review of Lees et al. (2010) in which driving simulators are postulated as ideal addition to on-road and cognitive testing. It represents a link between laboratory cognitive testing (high experimental control, low fidelity) and driving in reality (low experimental control, high fidelity). Age and mileage were used as covariates.

### 3. Results

Tables 5–7 show the average performance scores of the three measurement domains (on-road, cognitive and driving simulator assessment).

Pearson correlations between the three indices (Table 8) revealed no significant correlation between the on-road and the cognitive index. Significant correlations have been identified between the driving simulator and the on-road index ($r = .53$,
significant negative correlations were found between chronological age and the on-road index (r = −.29, p < .05), the cognitive index (r = −.54, p < .01) and the driving simulator index (r = −.46, p = .01).

Calculating a multiple linear regression with driving simulator index as dependent variable and on-road index as well as cognitive index as predictors while statistically controlling for age and mileages revealed a significant result explaining 43.1%
of the driving simulator-index variance by the two predictors (on-road and cognitive index). Even after exclusion of two participants (due to a traumatic accident during childhood and test anxiety) this model still explains 50.0% of the variance in the driving simulator index \( F(2,44) = 21.998, p < .001 \). Test conditions for the regression model were fulfilled (normal distribution, homoscedasticity, variance inflation factor (VIF = 1.03) for multicollinearity, Durbin-Watson test (value = 2.19) for autocorrelation) (Fig. 1).

In this regression model (Table 9) the on-road and the cognitive index are significant predictors of the driving simulator index (on-road index: \( t(44) = 4.96, p < .001, r = .59 \); cognitive index: \( t(44) = 3.58, p = .001, r = .47 \)). Participants with better on-road driving performance performed significantly better than participants with low on-road driving skills in the driving simulator. Correspondingly, participants with better cognitive abilities performed significantly better than participants with lower cognitive abilities in the driving simulator. Age and mileage did not exert statistically significant influences in this regression model (age: \( t(38) = 1.29, p = .20 \); sex: \( t(38) = -.71, p = .47 \); mileage: \( t(38) = -.73, p = .46 \)). Even in the model with the covariates (age, sex, mileage) the data fulfilled all prerequisites for interpreting the model. Test conditions for the regression model were fulfilled (normal distribution, homoscedasticity, variance inflation factor (1.56 ≤ VIF ≤ 5.53) for multicollinearity, Durbin-Watson test (value = 1.75) for autocorrelation).

4. Discussion

The aim of this paper was to examine the predictive value of older peoples driving simulator performance through cognitive performance and road test performance. We identified that driving simulator performance is statistically predicted (50% of explained variance) by performance measures obtained during on-road driving and cognitive testing independent from age. In the following we will discuss our findings in relation to the current literature on aging and driving performance.

Healthy older driver with a valid drivers' licence are a particular interesting sample because Siren and Meng (2012) showed that screening policies failed in identifying older unsafe drivers. Thus an interpolation from the driving behaviour of younger participants to older participants is most likely not valid, because driving behaviour and the type of car accidents is changing during aging (Braitman, Kirley, Ferguson, & Chaudhary, 2007). Older drivers differ from younger drivers in terms of cognitive, motor, and sensory functions but also in terms of driving experience, driving strategies and mileage (Anstey et al., 2005; Emerson et al., 2012). Moreover problematic driving behaviour in hazardous or complex traffic situations

Table 9

Results of the regression analysis with driving simulator performance as dependent variable and on-road test performance, cognitive test performance, chronological age, mileage, and sex as independent variables (n = 47).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>B (SD)</th>
<th>t</th>
<th>p-Value</th>
<th>r Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road test index</td>
<td>.537</td>
<td>.108</td>
<td>4.964</td>
<td>&lt;.001</td>
<td>.599</td>
</tr>
<tr>
<td>Cognitive index</td>
<td>.322</td>
<td>.090</td>
<td>3.575</td>
<td>.001</td>
<td>.474</td>
</tr>
<tr>
<td>Chronological age</td>
<td>.178</td>
<td>.138</td>
<td>1.292</td>
<td>.204</td>
<td>.205</td>
</tr>
<tr>
<td>Mileage</td>
<td>-.087</td>
<td>.121</td>
<td>-.717</td>
<td>.478</td>
<td>-.115</td>
</tr>
<tr>
<td>Sex</td>
<td>-.205</td>
<td>.277</td>
<td>-.739</td>
<td>.464</td>
<td>-.119</td>
</tr>
</tbody>
</table>

\( F(2, 44) = 21.998, P < .001, R^2 = .50 \)
(Horswill et al., 2008; Vance et al., 2006), compensational driving behaviour (Raw, Kountouriotis, Mon-Williams, & Wilkie, 2012) and a correlation between chronological age and cognitive decline (Borowsky et al., 2010) in older driver is well documented (Emerson et al., 2012).

A major point we are interested in is whether the driving performance in a driving simulator correlates with other tests which are used to examine driving performance in older participants. With a composite measure of cognitive performance (cognitive index) and the ratings of the on-road test (on-road index) we used two of the common tests as predictors to explain driving simulator performance. These two predictors explain 50% of the variance \((r = .71)\) in the driving simulator performance (simulator index). The driving simulator performance is positively related to both predictors, stronger however to the on-road index \((R^2 = 0.36)\) than to the cognitive index \((R^2 = 0.22)\). Nevertheless both predictors significantly explained substantial parts of the variance in the driving simulator data, what is in agreement with the neuroergonomic approach of the review of Lees et al. (2010).

During the driving simulator sessions we implemented normal and hazardous driving events. In line with the literature we expect in our on-road test behaviour of older drivers to contain compensational behaviour (for more detail see (Raw et al., 2012)), whereas our cognitive test performance is uninfluenced by compensational behaviour due to its novelty (highly correlated with chronological age, for more detail see (Borowsky et al., 2010). Therefore we suspect our driving simulator index to contain elements of normal on-road driving behaviour (compensation possible) as well as environmental elements in hazardous events (no compensation possible). During a driving career most drivers develop specific expertise to handle the car even during complex traffic situations. Thus drivers will skillfully drive in most of the standard and hazardous traffic situations by using automatic driving manoeuvres. These strategies are mostly procedural and not subject to conscious control (Sheridan, 2004). Bélanger et al. (2010) showed in their driving simulator study for older drivers that a reduction in automatic responses (car overtaking) correspond to reduced cognitive performance (selective and divided attention). Alternative compensatory behaviour (e.g., lower speed) wasn’t sufficient in hazardous traffic situation to prevent crashes. In addition, it is known that many older drivers are employing specific (explicit) strategies to avoid possible dangerous traffic situations (e.g., anticipatory speed reduction, avoidance of multi tasking during driving and avoidance of driving in unfamiliar areas) (Raw et al., 2012; Vance et al., 2006). Moreover there is evidence that older drivers show higher crash risk in complex traffic situations (intersection) compared to younger drivers (Braitman et al., 2007). In summary, the previous literature shows two sides of the same coin: It seems there are a reduction in automatic driving manoeuvres and an increase in explicit driving strategies without reducing crash risk in complex traffic situations. In our data we found no significant correlation between cognitive tests and on-road test. One probable reason of this result is that all on-road tests were completed without hazardous events hence we suspect that the drive to contain familiar situations only. Due to this lack of hazardous events for the DI it was not apparent if the driving performance of participants resulted from automatic or compensatory behaviour. In line with the literature concerning cognitive deficits it is essential to discriminate between driving performance based on automatic and compensatory behaviour which explains the weak correlation in our study (Selander et al., 2011). In this context we outline a study which compared healthy and demented older drivers in a driving simulator. The results showed little differences during baseline driving (except steering wheel position), but more high risk behaviours as a consequence of a hazardous event (unexpected stop of a lead vehicle) by demented drivers (Uc, Rizzo, Anderson, Shi, & Dawson, 2006). Our data support the interpretation that in our sample compensational driving strategies were used. As mentioned before we conclude that the driving simulator contains elements of normal on-road driving behaviour (compensation possible) as well as environmental elements in hazardous events (no compensation possible). Therefore we assume the relation between the on-road index and the driving simulator index to arise from normal driving behaviour (with compensatory behaviour) and the relation between the cognitive index and the driving simulator index to arise from behaviour in situations due to novelty (compensatory behaviour failed). Whether this speculation turns out to be valid has to be shown in future experiments.

Chronological age showed a weaker relation with the on-road index \((r = -.29)\) than with the driving simulator or the cognitive index \((r = -.46, r = -.54)\). These findings are in line with our interpretation which indicates that our on-road test is not capable to differentiate between automatic and compensational driving behaviour. The detailed results based on the single variables of each assessment are more closely linked to this interpretation. Single test variables of the cognitive assessment (except AMT) were highly correlated with chronological age, not so for single variables for the on-road and driving simulator assessment. For driving simulator variables we suspected reaction time to correlate with chronological age, because it was registered during the hazardous events. But only for the rural scenario there was a significant correlation between reaction time \((r = .42)\) and chronological age. No correlation was found for the urban street scenario. Perhaps the complexity of the urban street scenario was too unfamiliar for the participants motivating them to reduce their driving speed stronger than necessary, a behaviour which was substantiated by the observation that some participants drove with an unusual low speed. Therefore the hazardous simulator events no longer corresponded to a realistic hazardous moment. Similar problematic behaviour tendencies (strategically driving behaviour, lower speed) also were observed in the study of Bélanger et al., 2010. It seems that older drivers use self-regulation strategies while driving in unfamiliar exposure and in fact that is an efficient compensation strategy to prevent hazardous driving situations. Another finding was the positive correlations between all rural road and urban street variables, what indicates that driving behaviour problems exists in simple and complex surroundings. In that context a limitation of our study is that we generated one single “driving error” variable instead of differentiating between the number of crashes due to hazardous situations, crashes in normal traffic situations and other driving errors.
The statistical model we used here was uninfluenced by sex, mileage or chronological age. That mileage exerts no influence in our model seems to be astonishing since previous studies have shown that mileage is associated with reducing driving skills (Langford, Methorst, & Hakamies-Blomqvist, 2006). Mileage not having an effect in our model might be due to the fact that we have examined only participants from a relative homogeneous age group. In a recent prospective study Aksan et al., 2012 also report no influence of mileage on cognitive fitness and on-road driving performance. In our study all participants were active drivers and might have been specifically motivated. Chronological age remaining with no influence on the model (explained variance = 5%) is an important issue. Nevertheless age correlated with all measures on the basis of univariate correlations (on-road index, cognitive index, and driving simulator index), which is highly consistent with the current literature (Hakamies-Blomqvist, 1998; Lee et al., 2003; Owsley & McGwin, 2010). As shown above for single variables of the on-road and simulator assessment there are no tendencies that driving skills are related to age. But based on the on-road and simulator index there is evidence for age dependent reduction in driving skills. In contrast the cognitive assessment already shows on the basis of the single variables a relation to chronological age. So it seems that compensation strategies are difficult to measure on the basis of a single variable in familiar behaviour skills (in reality or in simulation). Nevertheless the overall performance seems to depend on ageing, so each single variable probably contains elements of incremental decreased performance, which is not visible in the correlations of single variables. For analysing compensational mechanism in older driver further research is needed. Our sample was too small and participants were highly motivated to participate, so that there is no final conclusion possible.

Interestingly, we did not find a significant relation between on-road test and cognition, which is different to Risser et al. (2008), who reported 50% explained variance in the on-road test. The reasons for the missing relation in our analysis are difficult to explain. However, there are several differences between our study and the study of Risser et al. (2008). First of all they have examined two participant samples (n = 204), which were younger (mean ages between 59 and 37.5 years) than the participants examined in our study (mean age 72 years). Secondly, Risser et al. (2008) worked with cut-off values (pass or fail) while we worked with continuous data for the on-road test. However, the composite measure we used for the cognitive performance is based on the sample and the algorithm used by Risser and colleagues probably does not match perfectly on our sample. It is possible that the laboratory situation while measuring cognitive performance is substantially unfamiliar (no compensatory skills available) compared to the familiar on-road test situation (automatic driving skills) especially for older participants. In this context findings in laboratory and field testing often uncovering different findings. In another study with a comparable setting the UFOV explained 20% of variance in the on-road test data in a sample of 85 participants (mean age 72 years). The authors argued that a normal driving error isn’t directly a dangerous situation (more a bad habit) and relates not automatically to a reduction in cognition. Also they found that participants who failed the on-road test were older than those who passed (Selander et al., 2011). As in their study the sample size of our study is rather small and there is a lack of inter-rater reliability of road test assessment. The subjective impression of the DI might be influenced by the bad habits of the drivers.

Our results support the findings of Lee et al., 2003. However, the influence of age on driving simulator and the on-road test was higher in their study than in ours. The reasons for these discrepancies are difficult to explain from the present point of knowledge. However, they have recruited a substantial larger sample of older drivers than we did (Lee et al. study: n = 129, our study: n = 49). Whether both samples differ in terms of IQ, socioeconomic status or other variables is currently difficult to assess. However, the sample of Lee et al. study included urban older driver residing in Perth while our sample consisted of older participants living in rural, suburban and urban districts around Zurich. Whether different living situations might exert substantial influences on cognition and driving behaviour have to be evaluating in future studies. In this case two interesting aspect should be reminded. Firstly in the context of mileage there is some evidence in differences of crash risk between rural and urban older driver (for more detail see Hanson & Hildebrand, 2011; Thompson, Baldock, Mathias, & Wundersitz, 2013). Secondly, recent studies have shown that a challenging and cognitively as well as socially demanding senior life is associated not only with emotional well-being but also with superior cognitive performance (Vergheese et al., 2003; Voelcker-Rehage, Godde, & Staudinger, 2011). A further difference to the Lee et al. study, which possibly might have substantial influence on the test results, is the fact that in their study an assessor has measured not only the on-road assessment but also half of the simulator assessment. Further all single variables in both assessments were correlated with age. So the assessments in the study of Lee et al. probably were not independent like in ours (VIF), which could not occur without inherent difficulties also by the aspect of objectivity.

The present study comprised a rather small sample and there is missing measures for inter-rater reliability. Furthermore the issue of SS should be raised. There is a possibility to handle SS because scores declined during study participation. Nevertheless a drop-out rate of 21% indicates problems in a large population subgroup.

5. Conclusion

This study reports that driving simulator performance is strongly related to on-road driving performance and (although to a lesser degree) to cognitive performance. Interestingly, age does not play a pivotal role in modulating the modelled relations. However, age is univariately related with the predictors (on-road driving, cognition) and the criterion (simulator driving). The close relation between on-road driving performance and performance in a simulator test qualifies sophisticated driving simulator tests for standard usage in evaluating driving performance in older participants.
Acknowledgements

This research project was funded by the University of Zurich Research Fund (Grant No. 56234109) and by Peter Ritter. We would like to thank Schuhfried GmbH for their Hard- and Software Support. We also thank Dr. Jacqueline Zoellig for fruitful discussions during the planning phase of this project. Thanks to all 21 trainees for their assistance with the data collection. Special thanks to Thomas Kniest, Benjamin Pertschy and Nadja Wohler for their encouragement during the study preparation and data acquisition.

References


