# Predicting On-Road Assessment Pass and Fail Outcomes in Older Drivers with Cognitive Impairment Using a Battery of Computerized Sensory-Motor and Cognitive Tests

*Petra A. Hoggarth, PhD,*†  Carrie R. H. Innes, PhD,*‡  John C. Dalrymple-Alford, PhD,*§¶ and Richard D. Jones, PhD*‡§¶***

**OBJECTIVES:** To generate a robust model of computerized sensory-motor and cognitive test performance to predict on-road driving assessment outcomes in older persons with diagnosed or suspected cognitive impairment.

**DESIGN:** A logistic regression model classified pass–fail outcomes of a blinded on-road driving assessment. Generalizability of the model was tested using leave-one-out cross-validation.

**SETTING:** Three specialist clinics in New Zealand.

**PARTICIPANTS:** Drivers (n = 279; mean age 78.4, 65% male) with diagnosed or suspected dementia, mild cognitive impairment, unspecified cognitive impairment, or memory problems referred for a medical driving assessment.

**MEASUREMENTS:** A computerized battery of sensory-motor and cognitive tests and an on-road medical driving assessment.

**RESULTS:** One hundred fifty-five participants (55.5%) received an on-road fail score. Binary logistic regression correctly classified 75.6% of the sample into on-road pass and fail groups. The cross-validation indicated accuracy of the model of 72.0% with sensitivity for detecting on-road fails of 73.5%, specificity of 70.2%, positive predictive value of 75.5%, and negative predictive value of 68%.

**CONCLUSION:** The off-road assessment prediction model resulted in a substantial number of people who were assessed as likely to fail despite passing an on-road assessment and vice versa. Thus, despite a large multicenter sample, the use of off-road tests previously found to be useful in other older populations, and a carefully constructed and tested prediction model, off-road measures have yet to be found that are sufficiently accurate to allow acceptable determination of on-road driving safety of cognitively impaired older drivers. **J Am Geriatr Soc 61:2192–2198, 2013.**

**Key words: driving; dementia; older adults; elderly**

From the *New Zealand Brain Research Institute, †Psychiatric Service for the Elderly, Princess Margaret Hospital, ‡Department of Medical Physics and Bioengineering, Christchurch Hospital, §Department of Psychology, University of Canterbury, ¶Department of Medicine, University of Otago, and **Departments of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand.

Address correspondence to Dr. Petra Hoggarth, New Zealand Brain Research Institute, 66 Stewart Street, Christchurch 8011, New Zealand. E-mail: petra.hoggarth@cdhb.govt.nz

DOI: 10.1111/jgs.12540

As the population of Western countries ages, a greater proportion of those driving will be aged 65 and older. More drivers will therefore have diseases of old age, especially cognitive impairment. Those with dementia have almost 2.5 times as many crashes that result in insurance claims as age-matched controls[1] and are 10.7 times as likely to be involved in a crash.[2] Nonetheless, a large proportion of people with early dementia are able to pass an on-road driving assessment, with observed pass rates ranging from 35% to 73%.[3–6]

Reviews of the dementia and driving literature have concluded that there is insufficient evidence to support the use of neuropsychological tests to determine driver safety in dementia.[7–9] The expense of an on-road assessment is currently needed to make decisions about on-road safety in this population, although more-recent work holds out hope that some tests may have predictive value. Previous studies[10,11] involving people with a wide variety of neurological disorders (70% stroke,[10] 33% possible or probable dementia, 31% stroke[11]) found that a combination of computerized sensory-motor and cognitive tests (*SMCTests*, Christchurch Neurotechnology Research Programme, Christchurch, New Zealand) could predict pass and fail outcomes on a medical driving assessment with 77% to 86% accuracy after leave-one-out cross-validation to estimate the prospective utility of the model. Similarly, another study[12] found that one *SMCTests* measure (Random tracking run 1) plus time to



0002-8614/13/$15.00

complete the Trail-Making Test B correctly classified 75% of on-road pass or fail outcomes after leave-one-out cross-validation in a sample of cognitively intact drivers aged 70 and older. The current study therefore investigated whether *SMCTests* would be useful in older adults with mild cognitive impairment (MCI) or suspected or diagnosed dementia.

Four studies previously employed classification models of pass–fail on-road driving test performance in those with MCI or suspected or diagnosed dementia. One study[13] recruited 115 individuals with MCI or dementia and used a maze completion task to classify pass and fail outcomes using binary logistic regression. It reported an accuracy of 77.4%, a sensitivity of 84.0%, and a specificity of 61.8%. Another study[3] recruited 99 people with dementia who completed various cognitive tests and the Washington University Road Test and reported an accuracy of 85%. A third study[6] studied 37 adults with dementia and a control group to compare the results of an extensive cognitive testing battery with those of an on-road driving assessment. Using discriminant analysis, the authors reported an impressive classification accuracy of 92.0%, a sensitivity of 90.0%, and a specificity of 93.0%. A fourth study[14] recruited 75 cognitively impaired drivers and administered a battery of physical and cognitive tests to predict whether they would pass or fail the on-road assessment. Using binary logistic regression, it found an accuracy of 71%, a sensitivity of 74%, and a specificity of 67%. Two of these studies tested the classification model using a validation procedure.[6,14] One[6] recruited a new sample of 17 participants and found that the accuracy of the model fell to 58.8% with a sensitivity of 40.0% and a specificity of 66.7%. The other[14] used a boot-strapping procedure and had a subsequent drop in overall accuracy from 71% to 57%. A number of factors, including sample size, differences between sample groups in the case of one study,[6] and overfitting of the original classification model, can explain these reductions in accuracy.

The current study constructed a logistic regression classification model of *SMCTests* measures for predicting pass and fail outcomes on an on-road driving assessment in a large sample of older adults. The model was tested for its ability to generalize to a new sample using leave-one-out cross-validation.

## METHODS

### Participants

Participants were 279 referrals (180 male, 99 female; mean age 78.4, range 56–92) to three New Zealand services that specialize in driving assessment of people with medical disorders. The Driving and Vehicle Assessment Service (DAVAS) at Burwood Hospital, Christchurch, assessed 155 people (101 male, 54 female); the Organisation of Therapy and Rehabilitation Services, Hamilton, assessed 76 people (43 male, 33 female); and Kevin O'Leary and Associates, Wellington, assessed 48 people (36 male, 12 female). All participants had diagnosed or suspected Alzheimer's disease, MCI, unspecified cognitive impairment, or memory problems. Participants had a current full drivers licence and had no need for driving adaptations on their vehicle. Participants gave informed consent, and the Upper South A Regional Ethics Committee, New Zealand, and the New Zealand Multi Region Ethics Committee, Wellington, New Zealand, approved the studies.

### Procedure

All participants completed an off-road battery of sensory-motor and cognitive tests (*SMCTests*) that used a car body with steering wheel and pedals (at the DAVAS service; Figure 1A) or a portable version (used by the other two services; Figure 1B). A subset from the *SMCTests* battery measuring arm reaction and movement times, visuomotor tracking, divided attention, complex attention, and planning was used. The Ballistic Movement test records reaction time, movement time, and peak velocity of arm movements when rapidly moving the steering wheel following a visually presented cue. The Sine and Random Tracking tests measure visuomotor coordination by recording the mean absolute error in millimeters of the tip of a vertically pointing arrow relative to a target when participants track two-dimensional sinusoidal and random targets (with 8-second previews), respectively, using the steering wheel. The Arrows Perception test requires participants to respond orally whether four simultaneously presented horizontal arrows are pointing in the same or different directions, with reaction time and number of correct responses recorded, measuring visual search speed and

**Figure 1.** (A) Modified car body used to run *SMCTests* at the Driving and Vehicle Assessment Service site. Test stimuli were projected onto a white wall in front of the apparatus. (B) Portable apparatus used at the other two sites. Stimuli for the portable version were presented on a computer monitor.

decision making ability. The Divided Attention test consists of concurrent testing of the Arrows Perception and Random Tracking tests, with participants asked to respond orally regarding arrow directions and to follow the random line target using the steering wheel. The Complex Attention test requires participants to move an arrow out of a box and across the screen using the steering wheel as quickly as possible following changing green-light stimuli. Recorded measures are reaction and movement times and lapses (when the arrow was not moved out of the box following the stimulus change) and invalid trials (when the arrow was not within the box when the stimulus changed). Complex Attention requires that participants focus on relevant cues and discount irrelevant cues and is sensitive to lapses in attention, which are detected according to invalid and lapse errors. The Planning test involves the presentation of a driving scene in plain view with participants instructed to "drive" the car along a road using the steering wheel, accelerator, and brake pedal. Obstacles to be negotiated include curves in the road, paint hazards, and intersections. Measures include number of paint hazards hit, number of collisions with other cars, safety margins between cars while crossing intersections, and number and duration of road position errors (including driving off the road). Performance requires the use of three types of apparatus (steering wheel, directional, and accelerator and brake pedal) to follow a series of rules in a unique environment (e.g., stopping at intersections and indicating appropriately while overtaking obstacles on the road while avoiding collisions with other vehicles).

More-detailed accounts of the specific tests, the measures derived from them, and their relationship to standard neuropsychological tests and to each other have been published previously.[12,15] Additional details are available in the *SMCTests* users manual (http://www.neurotech.org.nz/files/CanDAT_SMCTests_User_Manual.pdf).

Participants completed a medical driving assessment conducted by one of eight experienced driving occupational therapists (OTs). Assessors were blind to the results of off-road testing. Participants were able to use their own cars. As is standard practice in New Zealand medical driving assessments, there were no fixed driving route or scoring criteria for determining an on-road pass or fail outcome. The OTs made a qualitative judgment of pass or fail at the end of the assessment depending on whether they believed the identified medical factor (e.g., cognitive impairment) was unacceptably affecting the person's ability to drive safely.

### Data Analysis

All independent variables were on ratio scales, and although skewing on some measures was expected, the large sample size would allow for normalization of most distributions. Thus *t*-tests were used for pairwise comparisons of test performance according to pass and fail group. Selecting variables to offer to a model on the basis of their univariate association with the dependent variable is a common first step to overfitting a model to the idiosyncrasies of the sample and reducing generalization to the population.[16] It has also been suggested that a ratio of variables to participants of 1:10 to 1:15 will help to reduce overfitting risk, which can occur when too many variables are offered to the model.[16] The number of participants in the current study was sufficient that the ratio of variables to participants would be at most 1:10 and would not require selecting of variables according to their univariate association with the dependent variable (pass–fail on the on-road assessment).

Variables with high multicollinearity were excluded in preparing the model. Multicollinearity was examined using the collinearity diagnostics function in SPSS version 11.5.0 (SPSS, Inc., Chicago, IL). These diagnostics are independent of the relationship between variables and the dependent variable. The lower the reported tolerance value, the more correlated a measure is with one or more of the other variables. A guideline of a tolerance less than 0.20 has been suggested to detect multicollinearity.[17] Variables with the lowest tolerance were deleted individually and the analysis rerun until all independent variables had tolerance values greater than 0.20.

The model was built in SPSS using a backward elimination procedure and tested using leave-one-out cross-validation using a script written in MATLAB Version 7.10.0.499 (R2010a, The MathWorks, Inc., Natick, MA). Each case was removed from the sample and the model retrained on the remaining participants, with a prediction made for the excluded case.[18] This procedure was repeated for every case, and accuracy rates were averaged across all iterations. This procedure estimates how the model would perform given a new case. Using a cross-validation approach is supported by a previous study[19] that found that this approach provides better internal validation and ability to generalize to a new sample than models that are trained and then tested on a held-back sample.

## RESULTS

### On-Road Assessment

One hundred fifty-five of the 279 participants (55.5%) failed the on-road driving assessment, with no difference in fail rates between men and women (100 of 180 men failed, 55 of 99 women failed; Fisher exact test, two-tailed $P > .99$). The fail group was significantly older (mean age 80.2) than the pass group (mean age 76.2) (*t*-test, two-tailed, $z = -4.76$, $P < .001$).

### Pass and Fail Groups: Significant Differences and Effect Sizes

Comparisons of the test results of the pass and fail groups are shown in Table 1. Also shown are Cohen effect-size for rank-transformed variables, which was chosen because of the nonnormal distributions of some variables.[20] There were differences between the pass and fail groups in all but three *SMCTests* measures, with the fail group performing worse than the pass group.

Because of the significant effect of age, an analysis of covariance (ANCOVA) was performed with each variable controlled for age. Only six variables retained a $P < .05$ when age was controlled for: Arrows Perception omission of response, Divided Attention arrows correct and omission

Table 1. Comparison of On-Road Pass and Fail Groups Using *T*-Tests

| Test Measure | Pass Group, n = 124 | Fail Group, n = 155 | *T*-Test *P*-Value | Cohen-Type Effect Size[a] |
|---|---|---|---|---|
| Age, mean ± SD | 76.2 ± 7.9 | 80.2 ± 5.6 | <.001 | 0.52 |
| Ballistic movement test, ms, grand mean left and right arms ± SD | | | | |
| Reaction time | 445.6 ± 172.9 | 565.9 ± 281.3 | <.001 | 0.65 |
| Movement time | 289.8 ± 76.8 | 361.1 ± 121.2 | <.001 | 0.73 |
| Total reaction and movement times | 735.4 ± 211.9 | 927.0 ± 353.3 | <.001 | 0.79 |
| Peak velocity | 758.7 ± 190.3 | 661.3 ± 182.4 | <.001 | −0.54 |
| Tracking tests, mm, mean ± SD | | | | |
| Sine tracking run 1 error | 22.3 ± 12.2 | 27.5 ± 15.6 | .002 | 0.37 |
| Sine tracking run 2 error | 14.9 ± 9.3 | 20.0 ± 13.1 | <.001 | 0.49 |
| Random tracking run 1 error | 13.2 ± 7.2 | 16.7 ± 8.1 | <.001 | 0.51 |
| Random tracking run 2 error | 11.7 ± 7.0 | 16.0 ± 8.0 | <.001 | 0.65 |
| Arrows perception test, mean ± SD | | | | |
| Number of arrows correct | 11.5 ± 1.1 | 11.2 ± 1.6 | .06 | −0.18 |
| Omission of arrows response | 0.2 ± 0.6 | 0.5 ± 1.2 | .01 | 0.38 |
| Divided attention test, mean ± SD | | | | |
| Tracking error, mm | 12.2 ± 5.5 | 17.2 ± 7.8 | <.001 | 0.77 |
| Number of arrows correct | 10.9 ± 2.0 | 9.6 ± 2.9 | <.001 | −0.54 |
| Omission of arrows response | 0.7 ± 1.8 | 2.0 ± 2.7 | <.001 | 0.70 |
| Complex attention test | | | | |
| Reaction time, ms, mean ± SD | 662.8 ± 191.1 | 848.4 ± 377.3 | <.001 | 0.70 |
| Movement time, ms, mean ± SD | 453.0 ± 146.8 | 568.4 ± 244.8 | <.001 | 0.56 |
| Total reaction and movement times, ms, grand mean ± SD | 1,117.5 ± 303.9 | 1,398.7 ± 426.2 | <.001 | 0.76 |
| Movement time standard deviation, ms | 114.6 | 159.4 | .001 | 0.45 |
| Reaction time standard deviation, ms | 257.1 | 333.3 | <.001 | 0.56 |
| Number of lapse errors, mean ± SD | 1.7 ± 3.5 | 3.1 ± 4.3 | .003 | 0.49 |
| Number of invalid trials, mean ± SD | 0.6 ± 1.2 | 2.0 ± 3.2 | <.001 | 0.63 |
| Planning test, mean ± SD | | | | |
| Lateral road position error, mm | 3.4 ± 2.8 | 3.6 ± 1.6 | .41 | 0.33 |
| Duration of positional faults, s | 12.8 ± 13.0 | 24.9 ± 23.4 | <.001 | 0.70 |
| Distance traveled, m | 3.6 ± 0.7 | 3.6 ± 0.8 | .98 | 0.03 |
| Intersection safety margin, mm | 25.9 ± 19.2 | 18.0 ± 18.4 | <.001 | −0.44 |
| Number of hazards hit | 2.3 ± 1.4 | 3.0 ± 1.3 | <.001 | 0.52 |
| Number of crashes | 1.9 ± 2.5 | 2.9 ± 3.1 | .002 | 0.40 |

[a] The Cohen-type effect size is calculated using the mean ranks of pass and fail groups.[20] Positive effect sizes show that a higher score on the measure was related to greater likelihood of failing the on-road assessment, whereas negative effect sizes show that a lower score on the measure was related to less likelihood of failing the on-road assessment.

SD = standard deviation.

of arrows response, Complex Attention lapses and invalid trials, and Planning hazards hit.

Collinearity diagnostics were performed with the 27 variables. Five variables were deleted because they had tolerance values less than 0.2: Ballistic Movement mean total reaction and movement time, Random Tracking runs 1 and 2, Divided Attention omission of arrows response, and Complex Attention mean total reaction and movement time. The remaining 22 variables were offered to the model, for a variable to participant ratio of 1:13.

The model accepted eight measures: age, Ballistic Movement movement time, Ballistic Movement peak velocity, Divided Attention tracking error, Complex Attention reaction time, Complex Attention number of invalid trials, Complex Attention numbers of lapses, and Planning duration of positional faults. Standardized beta weights must be interpreted with caution because a 1 standard deviation–change in one measure may not be directly comparable with the same size change in another variable. Keeping this in mind, Table 2 shows that Ballistic Movement movement time and peak velocity measures were the two strongest predictors, followed by two Complex Attention measures, with age as the fifth strongest. These measures accounted for 36% of the variance in the on-road outcome (Nagelkerke coefficient of determination). The area under the receiver operating characteristic curve was 0.81 (95% confidence interval = 0.76–0.86). Using an a priori cut-point of 0.5 to determine pass or fail, the model correctly classified 211 of 279 participants (75.6%), with sensitivity for detecting fails of 78.7% and specificity of 71.8%.

After averaging of the 279 iterations that leave-one-out cross-validation generated, accuracy was estimated to be 72.0%, with a sensitivity of 73.5% and specificity of 70.2%. The positive predictive value after leave-one-out was 75.5%, meaning that, for all those predicted to fail the on-road assessment, the model was correct three-quarters of the time. The negative predictive power was 68.0%.

## DISCUSSION

A model with eight variables correctly classified 75.6% of participants with diagnosed or suspected MCI or dementia

Table 2. Variables Accepted into the Binary Logistic Regression Model

| Accepted Variable | B | B[a] | Wald | *P*-Value | Exponentiation of B Coefficient (Odds Ratio) |
|---|---|---|---|---|---|
| Ballistic movement, movement time | 0.013 | 0.338 | 9.842 | .002 | 1.013 |
| Ballistic movement, peak velocity | 0.004 | 0.187 | 6.215 | .01 | 1.004 |
| Complex attention, reaction time | 0.002 | 0.158 | 4.911 | .03 | 1.002 |
| Complex attention, invalid trials | 0.177 | 0.113 | 3.978 | .046 | 1.194 |
| Age | 0.062 | 0.107 | 8.083 | .004 | 1.064 |
| Divided attention, tracking error | 0.052 | 0.094 | 4.091 | .04 | 1.054 |
| Complex attention, lapses | −0.079 | −0.077 | 3.004 | .08 | 0.924 |
| Planning, duration of positional faults | 0.015 | 0.075 | 2.507 | .11 | 1.015 |
| Constant | −14.265 | | 19.231 | | |

[a] Standardized beta weights.[21]

with respect to on-road assessment pass and fail. ANCOVA showed that most of the simple effects shown in Table 1 became nonsignificant when age was entered as a covariate, with measures of reaction and movement times most affected. This may indicate that the measures of errors and lapses are more age independent. The model accepted age but also seven other variables, with age listed as the fifth most influential variable according to standardized beta weights. The measures accepted into the model show the importance of simpler sensory-motor tests such as Ballistic Movement but also those that rely on higher cognitive processes, such as the Complex Attention, Divided Attention, and Planning tests, which suggests that a number of factors are responsible for a drop in driving ability in this cognitively impaired sample and that there is unlikely to be a simple and quick-to-administer test that measures sufficient domains to produce a high accuracy of prediction of driving ability. The leave-one-out cross-validation produced small decrements in sensitivity and specificity, with a decline in overall accuracy of only 3.6%. Despite the large sample size, the use of off-road tests found to be useful in other populations, and a carefully constructed and tested prediction model, the authors concur with previous reviews that off-road measures that are sufficiently accurate to be used alone for determining the on-road driving safety of cognitively impaired older drivers have not been found.[7–9]

The results of previous studies[6,14] that found that the accuracy of their models dropped after a validation procedure (an accuracy drop of 92% to 58.8%[6] and 71% to 57%[14]) reinforce the importance of testing models beyond classification alone. These results show that initial classification models may be overfitted, which can reduce their generalization to the population. One study[6] did not report the number of variables offered to its model, only that 14 variables were accepted. Even if only the 14 variables accepted had been offered, the study would have had a ratio of 1 variable to 2.6 participants. This high ratio is likely to have resulted in an overfitted classification model. The other study[14] reported that six variables were accepted into its model, equating to a ratio of 1 variable to 12.5 participants, but the number of variables offered to the model is not reported, so this ratio cannot be determined to consider the possibility of overfitting accounting for the drop in accuracy. The study[14] stated that the accuracy of its model using basic physical tests and cognitive measures (Trail-Making Test B, computerized maze error score, Mini-Mental State Examination, and clock drawing) was not high enough to be used as a screening test and that more-accurate measures were required.

The current study produced a model that was higher in post-validation accuracy than those studies[6,14] but less accurate than others constructed using *SMCTests* measures in healthy older drivers (75% accuracy)[12] and in those with brain disorders (77–86% accuracy).[10,11] The healthy older drivers study[12] accepted a single *SMCTests* measure, Random Tracking run 1, that was not accepted into the current model. Three measures accepted into the model for one of the brain disorders study (n = 501)[11] were the same as in the current study (Divided Attention tracking error, Complex Attention mean reaction time, and age), although the brain disorders study had some participants in common with the current study, meaning that the samples were not independent. The other brain disorders study (n = 50)[10] shared Divided Attention tracking error in common with the current study and the first brain disorders study.[11] The fact that all three studies shared the Divided Attention tracking error score may indicate that decrement of accuracy in a divided attention task is a useful off-road measure of driving ability regardless of specific brain pathology.

The off-road testing in the current study had unacceptably high levels of false-positive and false-negative errors, and removal of on-road assessment is not warranted at this stage, although the subset of eight tests identified here could be useful in a screening capacity to reduce the number of people referred for an on-road assessment. If a cut-point of 0.40 was chosen (rather the default of 0.50 reported above), then the model following leave-one cross-validation produced a sensitivity of 80%, a specificity of 62.1%, and an overall accuracy of 72.0%. By identifying those most likely to fail, this screening cut-point would reduce the number of on-road assessments required by approximately 40%, although at this cut-point, 20% of those who would fail an on-road assessment would pass the screen and not be required to take any on-road assessment. Whether this level of sensitivity for identifying people who would fail an on-road assessment is acceptable depends on the resources of the specific driver licensing system and current rate of on-road testing. It is informative to compare the base-rate of adverse driving outcomes, such as crashes, of drivers with dementia with those of healthy controls. One study[1] found that drivers with dementia had an average of one crash every 6.5 years, compared with a randomized stratified sample of controls, who had one crash every 16.5 years. Whether this level of crash risk for the false negatives missed by a screening test is acceptable could contribute to discussion and policy concerning road safety.

## Limitations

Although their general practitioners or a memory assessment clinic referred all 279 participants, they did not all have a specific diagnosis of MCI or dementia. A subset of 32 participants completed additional neuropsychological evaluation and interviews with significant others (by author PAH), which confirmed their specific diagnosis of probable Alzheimer's disease (n = 24) or MCI (n = 8). One participant was found to have neither dementia nor MCI and was excluded from the study. If this subset was representative of the sample as a whole, it could be expected that approximately 75% of the sample would be diagnosed with dementia. Lack of knowledge of the diagnoses of the entire sample reduces the model's ability to generalize to a pure sample of participants with diagnosed dementia.

The nature of on-road assessment may be another limitation. The current study followed standard practice for medical driving assessments in New Zealand and did not use a standardized scoring system for determination of pass and fail outcomes. Such nonstandardized assessments are common practice. A survey of 114 U.S. and Canadian driving assessors found that only 24% of assessors used a standardized scoring system for on-road assessments and that only 10% used a cutoff score to assess driving competency.[22] A subset of 60 participants from the current study had additional information collected in the form of a 29-item driving behavior checklist completed by the driving OT.[23] Eighteen items were significantly associated with pass and fail outcomes, with the two most frequently rated reasons for failing being decreased awareness of other road users and decreased awareness of environment.

It would be useful to know whether those who failed an on-road assessment had a greater number of adverse driving outcomes than those who passed. Prospective studies are not possible in a clinical sample, in which the outcome of the on-road assessment is licence revocation. A recent study[24] found no significantly greater rates of prospectively measured self- and state-reported car crashes and traffic infringements in healthy older drivers who failed an on-road assessment than in those who passed. With regard to retrospectively reported crashes, the nature of a degenerative cognitive disorder reduces the validity of these data to reflect future driving behavior. As such, these data were not collected for the current study.

## ACKNOWLEDGMENTS

**Conflict of Interest:** Funding for this project was received from the Canterbury Medical Research Foundation, the Accident Compensation Corporation, Road Safety Trust. The editor in chief has reviewed the conflict of interest checklist provided by the authors and has determined that the authors have no financial or any other kind of personal conflicts with this paper.

**Author Contributions:** Hoggarth: conception and design of study, off-road testing of 60 participants, analysis and interpretation of data, primary writer of article. Innes: conception and design of study, collation of data for 219 participants, analysis and interpretation of data, review and revision of article to final stage. Dalrymple-Alford: conception and design of study, analysis design and interpretation, review and revision of article to final stage. Jones: conception and design of study, analysis design and interpretation, review and revision of article to final stage.

**Sponsor's Role:** None of the sponsors of this research played any role in the design; methods; participant recruitment; data collection; analysis or interpretation of data; or preparation, review, or approval of manuscript.

## REFERENCES

1. Cooper PJ, Tallman K, Tuokko H et al. Vehicle crash involvement and cognitive deficit in older drivers. J Safety Res 1993;24:9–17.
2. Zuin D, Ortiz J, Lopez OL. Motor vehicle crashes and abnormal driving behaviours in patients with dementia in Mendoza, Argentina. Eur J Neurol 2002;9:29–34.
3. Carr DB, Barco PP, Wallendorf MJ et al. Predicting road test performance in drivers with dementia. J Am Geriatr Soc 2011;59:2112–2117.
4. Fox GK, Bowden SC, Bashford GM et al. Alzheimer's disease and driving: Prediction and assessment or driving performance. J Am Geriatr Soc 1997;45:949–953.
5. Grace J, Amik MM, D'Abreu A et al. Neuropsychological deficits associated with driving performance in Parkinson's and Alzheimer's disease. J Int Neuropsychol Soc 2005;11:766–775.
6. Lincoln NB, Radford KA, Lee E et al. The assessment of fitness to drive in people with dementia. Int J Geriatr Psychiatry 2006;21:1044–1051.
7. Australian and New Zealand Society for Geriatric Medicine. Position Statement: Dementia and Driving. Australas J Ageing 2010;29:137–141.
8. Marino M, de Belvis A, Basso D et al. Interventions to evaluate fitness to drive among people with chronic conditions: Systematic review of literature. Accid Anal Prev 2013;50:377–396.
9. Iverson DJ, Gronseth GS, Reger MA et al. Practice parameter update: Evaluations and management of driving risk in dementia: Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2010;74:1316–1324.
10. Innes CRH, Jones RD, Dalrymple-Alford JC et al. Sensory-motor and cognitive tests can predict driving ability of persons with brain disorders. J Neurol Sci 2007;260:188–198.
11. Innes CRH, Lee D, Chen C et al. Do complex models increase prediction of complex behaviours? Predicting driving ability in people with brain disorders. Q J Exp Psychol (Hove) 2011;64:1714–1725.
12. Hoggarth PA, Innes CRH, Dalrymple-Alford JC et al. Comparison of a linear and a non-linear model for using sensory-motor, cognitive, personality, and demographic data to predict driving ability in healthy older adults. Accid Anal Prev 2010;42:1759–1768.
13. Snellgrove CA. Cognitive Screening for the Safe Driving Competence of Older People with Mild Cognitive Impairment or Early Dementia. Canberra: Australian Transport Safety Bureau, 2000.
14. Ott BR, Davis JD, Papandonatos GD et al. Assessment of driving-related skills prediction of unsafe driving in older adults in the office setting. J Am Geriatr Soc 2013;61:1164–1169.
15. Innes CRH, Jones RD, Anderson TJ et al. Performance in normal subjects on a novel battery of driving-related sensory-motor and cognitive tests. Behav Res Methods 2009;41:284–294.
16. Babyak MA. What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. Psychosom Med 2004;66:411–421.
17. Garson GD. Multiple Regression, 2010 [on-line]. Available at http://faculty.chass.ncsu.edu/garson/PA765/regress.htm Accessed January 6, 2010.
18. Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Francisco: Morgan Kaufmann, 2000.

19. Steyerberg EW, Harrell FE, Borsboom GJJM et al. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54:774–781.
20. Hopkins WG. A New View of Statistics, 2004 [on-line]. Available at www.sportsci.org/resource/stats/ Accessed August 16, 2004.
21. King JE. Standardized Coefficients in Logistic Regression, 2007 [on-line]. Available at http://www.ccitonline.org/jking/homepage/ Accessed June 22, 2013.
22. Korner-Bitensky N, Bitensky J, Sofer S et al. Driving evaluation practices of clinicians working in the United States and Canada. Am J Occup Ther 2006;60:428–434.
23. Hoggarth P, Innes C, Dalrymple-Alford J et al. On-road driving assessment errors associated with pass and fail outcomes for older drivers with cognitive impairment. Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Lake Tahoe, California, 2011.
24. Hoggarth PA, Innes CRH, Dalrymple-Alford JC et al. Prospective study of healthy older drivers: No increase in crash involvement or traffic citations at 24 months following a failed on-road assessment. Transp Res F Traffic Psychol Behav 2013;16:73–80.