

Accepted Manuscript

Title: Can we improve clinical prediction of at-risk older drivers?

Author: Alex R. Bowers R. Julius Anastasio Sarah S. Sheldon
Margaret G. O'Connor Ann M. Hollis Piers D. Howe Todd S.
Horowitz



PII: S0001-4575(13)00265-0
DOI: <http://dx.doi.org/doi:10.1016/j.aap.2013.06.037>
Reference: AAP 3214

To appear in: *Accident Analysis and Prevention*

Received date: 8-11-2012
Revised date: 28-6-2013
Accepted date: 30-6-2013

Please cite this article as: Bowers, A.R., Anastasio, R.J., Sheldon, S.S., O'Connor, M.G., Hollis, A.M., Howe, P.D., Horowitz, T.S., Can we improve clinical prediction of at-risk older drivers?, *Accident Analysis and Prevention* (2013), <http://dx.doi.org/10.1016/j.aap.2013.06.037>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Running head: PREDICTING AT-RISK OLDER DRIVERS

Can we improve clinical prediction of at-risk older drivers?

Alex R. Bowers^{a,b} PhD, R. Julius Anastasio^a MS, Sarah S. Sheldon^a BS,
Margaret G. O'Connor^{c,d} PhD, Ann M. Hollis^c OTR/L, Piers D. Howe^e PhD,
Todd S. Horowitz^{b,f} PhD

^a Schepens Eye Research Institute, Massachusetts Eye and Ear, Boston, MA, USA

^b Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

^c Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, USA

^d Department of Neurology, Harvard Medical School, Boston, MA, USA

^e Psychological Sciences, University of Melbourne, Melbourne, VIC, Australia

^f Department of Surgery, Brigham and Women's Hospital, Boston, MA, USA

Corresponding author:

Alex Bowers

The Schepens Eye Research Institute

20 Staniford Street

Boston MA 02114

USA

Phone: 1-617-912-2512

Fax: 1-617-912-0112

Email: alex_bowers@meei.harvard.edu

Email addresses of other authors:

julius.anastasio@gmail.com

sarah.sheldon@schepens.harvard.edu

moconnor@bidmc.harvard.edu

pdhowe@unimelb.edu.au

ahollis@bidmc.harvard.edu

todd.horowitz@nih.gov

Abstract

Objectives: To conduct a pilot study to evaluate the predictive value of the Montreal Cognitive Assessment test (MoCA) and a brief test of multiple object tracking (MOT) relative to other tests of cognition and attention in identifying at-risk older drivers, and to determine which combination of tests provided the best overall prediction.

Methods: Forty-seven currently-licensed drivers (58 to 95 years), primarily from a clinical driving evaluation program, participated. Their performance was measured on: (1) a screening test battery, comprising MoCA, MOT, MiniMental State Examination (MMSE), Trail-Making Test, visual acuity, contrast sensitivity, and Useful Field of View (UFOV); and (2) a standardized road test.

Results: Eighteen participants were rated at-risk on the road test. UFOV subtest 2 was the best single predictor with an area under the curve (AUC) of .84. Neither MoCA nor MOT was a better predictor of the at-risk outcome than either MMSE or UFOV, respectively. The best four-test combination (MMSE, UFOV subtest 2, visual acuity and contrast sensitivity) was able to identify at-risk drivers with 95% specificity and 80% sensitivity (.91 AUC).

Conclusions: Although the best four-test combination was much better than a single test in identifying at-risk drivers, there is still much work to do in this field to establish test batteries that have both high sensitivity and specificity.

Key words: driving safety; multiple object tracking; visual attention

1. INTRODUCTION

The number of older drivers is growing rapidly. In 2009 there were 7.7 million drivers \geq 80 years in the U.S. (Federal Highway Administration Department of Transportation (US) 2009); a 47% increase compared to 1999 (Federal Highway Administration Department of Transportation (US) 1999). Drivers in this age group are at an elevated risk for accidents relative to middle-aged drivers (McGwin and Brown 1999) and are more likely to be fatally injured (Lyman *et al.* 2002). However, it is not appropriate to simply prohibit people from driving on the basis of chronological age. For many older people, driving is important for independence and quality of life; indeed, driving cessation is linked with social isolation and depression (Marottoli *et al.* 1997, Fonda *et al.* 2001, Edwards *et al.* 2009). Thus, it is important to be able to accurately distinguish between older drivers who are safe to continue driving, and those who might be at-risk and should cease driving. Unfortunately, this is not a straightforward problem.

Since driving is a complex task, a combination of multiple tests may be more likely to predict driver performance than any single test (Wood *et al.* 2008). Failures in sensory, cognitive, or motor abilities with increasing age could all contribute to driving failures, and no one test would be likely to capture all these aspects (Anstey *et al.* 2005). This is the approach adopted by clinical driver evaluation programs, which typically include a battery of screening tests and an on-road driving test (Korner-Bitensky *et al.* 2006). A growing number of test batteries have been proposed and examined (Hoffman *et al.* 2005, Oswanski *et al.* 2007, Bédard *et al.* 2008, Wood *et al.* 2008, Kay *et al.* 2009, Korner-Bitensky and Sofer 2009, Dobbs and Schopflocher 2010, Carr *et al.* 2011, Wood *et al.* 2013). However, as yet, none provide sufficiently good sensitivity *and* specificity either for mass screening of older drivers or to be a replacement for an on-road test (Bédard *et al.* 2008, Kay *et al.* 2012). Therefore, at the moment,

screening batteries in driver evaluation programs are mainly used to provide information to supplement the road test, and possibly identify drivers for whom an on-road test would be unsafe.

A recent review suggested that a screening battery, as a replacement for a road test, should achieve both sensitivity and specificity of at least 90% (Kay *et al.* 2012); however, none of the batteries tested to date have reached that goal for a binary classification of safe vs. at-risk drivers (Table 1). For example, although a multi-disciplinary battery including vision, cognitive and motor performance tests evaluated in a non-clinical population was relatively good at identifying at-risk drivers (91% sensitivity), 30% of safe drivers were incorrectly categorized as being unsafe (70% specificity) (Wood *et al.* 2008). On the other hand, in clinical populations (people referred to a driving assessment program), the DriveAble screen battery was relatively good at identifying safe drivers (specificity 90%), but failed to identify almost one-quarter of at-risk drivers (sensitivity 76%) (Korner-Bitensky and Sofer 2009), while the DriveSafe/DriveAware battery (Kay *et al.* 2009) and the SIMARD battery (Dobbs and Schopflocher 2010) both achieved high sensitivity (97% and 93%, respectively), but lower specificity (58% and 40%) for a binary classification (Table 1).

These findings underscore the importance of continuing to evaluate individual tests and combinations of tests with the aim of achieving both high sensitivity and high specificity with as few tests as possible. One approach to developing such a battery would be to incorporate tests that precisely target different functions that are both critical to driving and sensitive to aging (and accompanying medical conditions). In this study we examined the predictive ability of two such tests that had not, to our knowledge, previously been evaluated as predictors of at-risk older drivers.

The first test was the Montreal Cognitive Assessment (MoCA, (Nasreddine *et al.* 2005)), which is a cognitive screening task similar in design to the Mini Mental State Exam (MMSE), but with additional subtests focusing on multi-tasking aspects of attention relevant to driving. It is also more sensitive to mild cognitive decline than the MMSE (Nazem *et al.* 2009, Freitas *et al.* 2013). Thus our hypothesis was that the MoCA might be a better predictor of on-road driving than the MMSE. The other test, Multiple Object Tracking (MOT; Pylyshyn and Storm 1988), is a computerized measure of visual attention, like the well-established Useful Field of View (UFOV, (Ball *et al.* 1988)). However, while the UFOV involves brief (<500 ms) presentations of static stimuli, MOT requires continuous attention to multiple moving objects over several seconds. Our hypothesis was that the sustained, dynamic nature of the task captures cognitive skills important for driving (Kunar *et al.* 2008) and may provide additional information about sustained attentional capabilities relevant to driving.

A cohort of older drivers underwent a comprehensive evaluation comprising a road test and a standard clinical cognitive assessment battery (including the MMSE and the Trail-Making Test) as used by DriveWise, a clinical driving assessment program (O'Connor *et al.* 2008). In addition, they completed the MoCA test, a brief MOT test developed for clinical populations (Bowers *et al.* 2011) and the UFOV (as a comparison for the MOT). We had three primary goals: 1) determining whether the MoCA and MOT provided new information regarding critical aspects of the cognitive abilities needed for safe driving; 2) determining whether adding MoCA and/or MOT and/or UFOV improved the predictive value of the standard clinical cognitive assessment battery; and 3) determining the combination of tests that provided the best overall prediction of the road test outcome. The study was conducted as a pilot in preparation for a future, larger sample study.

2. METHODS

2.1. Participants

As this was a pilot study, we recruited a convenience sample of 32 consecutive participants from DriveWise, a clinical driving assessment program at Beth Israel Deaconess Medical Center to which people are referred if there is a concern about whether or not they should be driving (O'Connor *et al.* 2008). Only DriveWise clients who were eligible for inclusion in the study were invited to participate. In addition, 15 older volunteers (with normal cognition) were included; they had previously participated in studies at Schepens Eye Research Institute, mostly as normally-sighted controls in driving simulator studies (Bronstad *et al.* 2013). Inclusion criteria were: a current valid driver's license, vision meeting the requirements for licensure in MA (visual acuity of at least 20/40 and visual field of at least 120° horizontal diameter), and no physical impairments that would limit interaction with a touch screen.

The age and sex distributions were similar for the two recruitment sources, but the proportion with mild cognitive impairment was higher in the DriveWise group (DriveWise 34%; Schepens 0%). Mild cognitive impairment was diagnosed by a cognitive neurologist or neuropsychologist using the Petersen (2004) criteria. Education data were available for 26 participants in the DriveWise group; of these, 22 had more than high-school education, two had 12 years, one had 10 years and one had 9 years education. All participants in the Schepens group had more than high-school education.

The study adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Boards (IRB) at Beth Israel and Schepens. All participants provided written informed consent.

2.2. Test battery

All participants were administered a test battery comprising: vision measures; the standard DriveWise clinical cognitive test battery (O'Connor *et al.* 2008); the MoCA; and two visual attention tests, the UFOV and a brief MOT test, both presented on a touch-screen monitor. Habitual eye glasses were worn for all tests.

The test battery took about 1 to 1.5 hours to complete, with breaks, as needed. At DriveWise, the tests were administered within the clinic schedule. Therefore, due to time limitations (such as participants arriving late for their appointment but having to start the on-road test on time), not all of the DriveWise participants were able to complete all tests; however, all Schepens participants completed all tests. The tests were administered in a standardized order at Schepens by author RJA (all tests), and at DriveWise, by RJA (MOT, and MoCA) and author AMH (vision tests, MMSE, Trails and UFOV).

2.2.1 Vision measures

Binocular visual acuity was measured using an ETDRS acuity chart, either freestanding or computerized, with each letter scored as 0.02 log units (higher scores indicating worse performance). Binocular letter contrast sensitivity was measured using a MARS chart (Arditi 2005), with each letter scored as 0.05 log units (higher scores indicating better performance). The MARS chart has good repeatability and is comparable to the well-established Pelli-Robson letter contrast sensitivity chart (Dougherty *et al.* 2005). Contrast sensitivity was measured as deficits in this aspect of visual function have been associated with mild cognitive impairment (Risacher *et al.* 2013), and it may be a better predictor of driving performance than visual acuity (Owsley and McGwin 2010).

2.2.2 DriveWise clinical cognitive test battery

The clinical cognitive test battery included the MMSE (Folstein *et al.* 1975) and the Trail-Making Tests (parts A and B) (Reitan 1955). The MMSE is a brief test commonly used to screen for cognitive impairment (range 0–30, higher scores indicating better functioning), which has been reported to be predictive of the road-test outcome in some, but not all, studies (Anstey *et al.* 2005, Bédard *et al.* 2008, Asimakopulos *et al.* 2012, Crizzle *et al.* 2012). The Trail-Making Test is a popular neuropsychological test that consists of two parts. In Part A, numbers are connected in ascending order. In Part B numbers and letters are connected in alternating sequential order (e.g., 1, A, 2, B, 3, C, etc.). The score on each part is the time (in seconds) required to complete the task (higher scores indicating worse functioning). The Trail-Making Test is widely used in driver assessment programs (Korner-Bitensky *et al.* 2006) and has been associated with the road test outcome in some, but not all, studies (Anstey *et al.* 2005, Bédard *et al.* 2008, Asimakopulos *et al.* 2012). In two large population-based studies, Trails B was predictive of at-fault motor vehicle collisions (Ball *et al.* 2006, Friedman *et al.* 2013).

2.2.3 Montreal Cognitive Assessment (MoCA)

The MoCA is a relatively new cognitive screening instrument which was designed to address some of the limitations of the MMSE (Nasreddine *et al.* 2005). It assesses a broader range of cognitive domains, contains additional subtests focusing on multi-tasking (including an alternation task adapted from Trails B) and is more sensitive to mild cognitive decline than the MMSE (Nazem *et al.* 2009, Freitas *et al.* 2013). The range of scores is 0 to 30, with higher scores representing better functioning.

2.2.4 Useful field of view (UFOV)

Processing speed and static visual attention were assessed using Subtests 1 and 2 from the commercially-available, PC-based UFOV test, version 6.0.9 (Edwards *et al.* 2005, Edwards *et al.* 2006). The first subtest (processing speed) required the identification of a central target (outline of a car or a truck). The second (divided attention) required identification of the central target, as well as localization of a peripheral target (car) presented simultaneously at one of eight radial locations about 11 cm from the center of the screen (45 cm viewing distance). Targets were displayed from 17 to 500 ms using a double staircase method, and the score for each subtest was expressed as the minimum display duration (in seconds) for which the subject achieved a 75% correct response rate (with longer durations representing poorer performance). UFOV has been demonstrated to be predictive of on-road driving performance and crashes in a number of studies (Clay *et al.* 2005), in particular UFOV subtest 2 (Owsley *et al.* 1998, Ball *et al.* 2006).

2.2.5 Brief multiple object tracking (MOT)

Sustained dynamic attention was evaluated with a brief MOT task (Bowers *et al.* 2011) in which six high contrast black disks subtending 2° visual angle were presented. At the beginning of each trial, three target disks were designated in green. All disks then turned black (so targets and distractors appeared identical) and moved randomly for 5-8 s. Participants were asked to track the three targets as they moved. At the end of the tracking period, the disks stopped and participants indicated which of the six were targets. Speeds were adjusted on each trial using a simple one-up, one-down staircase. The speed threshold for 60% correct performance was determined using the QUEST algorithm (King-Smith *et al.* 1994), with higher scores representing better performance. Participants underwent 10 practice trials of the brief MOT task before completing 50 experimental trials.

2.3. On-road evaluation

Each participant underwent a 45-minute (7.5 mile) standardized road test based on the Washington University Road Test (Hunt *et al.* 1997), including urban, local and highway roads. All tests were conducted between 1:00 and 3:00 pm by an occupational therapist (OT), author AMH, and a certified driving rehabilitation instructor (DI) in a dual-control car. The DI provided directions and monitored overall safety, while the OT sat in the back seat and recorded behavioral performance for a range of maneuvers along the route. The DI was masked to all scores on the test battery. The OT was also masked to all scores for Schepens participants, but was only masked to scores on the MoCA and MOT tests for DriveWise participants.

A standardized set of items was scored within each maneuver (a total of 93 items along the whole route). For non-highway driving, items were grouped into seven categories (Wood *et al.* 2009): Mirrors (checks blind spot, rear-view mirror), gap selection (appropriate gap), indication (signals), lane position (lane keeping), observation (traffic light observance, checks traffic, performs maneuver when safe), planning (smoothness of lane change, does not hesitate without reason before proceeding, plans for stop by slowing vehicle), and speed control. Highway driving was included as a separate category. Each item was rated “pass” or “error”. Verbal prompts and physical interventions were recorded and classified as errors. For each participant, the total number of error items was computed. Error totals were expressed as a proportion of the total number of items within a given category.

The OT and DI independently used a three-point rating scale (0 = pass, 1 = marginal, 2 = fail) to assign a global judgment of driving safety based on performance along the whole route, including consideration of physical and/or verbal interventions. In practice, raters occasionally indicated scores in between 0 and 1, or 1 and 2, which were coded as 0.5 and 1.5, respectively.

Inter-rater agreement was good, with sixty-nine percent of OT and DI global ratings being the same, and an intraclass correlation coefficient of 0.82 ($F(47, 48) = 10, p < .001$; 95% bootstrap CI = 0.70-0.89). For analysis purposes, the two global ratings were averaged and then coded as “at-risk” (average global rating ≥ 0.5 ; i.e. at least one of the raters gave a marginal score) or “safe” (average global rating < 0.5).

2.4. Data analyses

Data were analysed using SPSS version 11.5 (Chicago IL, USA) and R version 2.14.0 (Kundu *et al.* 2011, R Development Core Team 2011, Robin *et al.* 2011). To determine the ability of each test to predict the at-risk/safe driving outcome, we computed receiver operating characteristic (ROC) curves and derived the area under the curve (AUC; ranges from 0.5 for chance performance to 1.0 for perfect prediction). Optimal cutoff points were identified using the Youden index, the threshold value for which sensitivity + specificity is maximized (Fluss *et al.* 2005).

To determine which combination of tests provided the best overall prediction, we fit data for each combination with a logistic regression model and then compared the models' performance with the integrated discrimination improvement (IDI) index (Pencina *et al.* 2008, Van Calster and Van Huffel 2010). The IDI can be seen as the difference between improvement in average sensitivity and any decrease in specificity. For each model, we also performed the Hosmer-Lemeshow test for goodness of fit, calculated the generalized R^2 (i.e., Nagelkerke R^2) and the AUC, and identified the optimal cutoff points of their ROC curve.

3. RESULTS

Twenty-nine participants were rated "safe" and eighteen "at-risk"; all of the at-risk participants were in the DriveWise group.

3.1. Test battery performance

Safe and at-risk participants did not differ in age, sex or vision measures (Table 2). However, a higher proportion of at-risk than safe participants had mild cognitive impairment (44% and 10%, respectively; Table 2). In fact, of the 11 participants with mild cognitive impairment, 8 (73%) were rated as being at-risk. At-risk participants had significantly worse performance on the MMSE, MoCA, Trails B and UFOV subtest 2 (Table 2), in part because participants with mild cognitive impairment performed significantly less well on these tests than non-impaired participants (Appendix, Table A.1).

While there were significant interrelationships among the cognitive and attention tests, we found no significant relationship between age and performance on the attention tasks or the MMSE. There was a significant, though weak correlation between age and performance on the MoCA (Table 3).

3.2. Driving behaviors

As expected, participants who were considered at-risk committed a greater proportion of driving errors than participants who were considered safe. In particular, at-risk participants had significantly more highway, observation, planning, speed control, and indication errors than safe participants (Figure 1). Furthermore, at-risk participants received significantly more interventions ($Mdn = 2.00$) than the safe group ($Mdn = .00$), $z = -4.44$, $p < .001$; Wilcoxon rank-sum test.

3.3. Predicting the driving test outcome from screening tests

For these analyses, participants with missing data were excluded on a test-by-test basis (Table 4); a similar pattern of results was obtained when only those participants who had completed all tests were included (Appendix, Table A.2).

Across the demographic, vision, cognitive, and attention measures, the UFOV subtest 2 had the greatest AUC, significantly greater than all other tests ($z = 2.02$ to 3.51 , $p = .043$ to $p = .004$) except for MoCA, MMSE and Trails B (Table 4 and Figure 2). MoCA had the second greatest AUC, significantly greater than that for visual acuity ($z = 1.97$, $p = .049$), but not significantly greater than the AUCs for the other measures (including MMSE). While the UFOV subtest 2 performed the best overall in predicting the driving test outcome, it was exceeded in specificity by both Trails A and the brief MOT. In sensitivity, the UFOV subtest 2 and contrast sensitivity were the best, followed by the MoCA (Table 4).

3.4. Combined predictors

Only participants who had completed all screening tests ($n = 32$; 10 at-risk and 22 safe) were included in these analyses (similar results were obtained when there were different numbers of subjects in each model; Appendix, Table A.3). Given that our overall battery had nine tests, we wished to avoid testing each possible combination. Instead, we started with a "default" baseline battery, the Clinical Model, and sought to improve on it by adding tests not already in the model. The Clinical Model included four tests typically administered in DriveWise evaluations: visual acuity, MMSE, and Trails A and B. This combination resulted in a model with an overall AUC of .79 (Table 5).

We first evaluated whether either MoCA or MOT improved this model. However, neither replacing MMSE by MoCA (AUC = .73; IDI [95% CI] = -.06 [-.16 - .03], $p = .20$), nor adding

MOT to the model (AUC = .79; IDI [95% CI] = .00 [-.02 - .02], $p = .99$) significantly improved the predictive power; thus MOT and MoCA were not considered further.

Next we looked at the effect of adding other tests not already in the Clinical Model, including, contrast sensitivity, the UFOV subtest 1 and UFOV subtest 2. Adding UFOV subtest 2 significantly improved model performance (AUC = .87; IDI [95% CI]: .15 [.02-.28]; $p = .02$). Contrast sensitivity was the next best addition (AUC = .82; IDI [95% CI]: .06 [-.02-.15]; $p = .14$). However, adding UFOV subtest 1 had no effect (AUC = .76; IDI [95% CI]: .05 [-.05-.15]; $p = .31$). Thus, UFOV subtest 2 and contrast sensitivity were added to our original Clinical Model which significantly improved its performance (AUC = .90; IDI [95% CI]: .23 [.08-.38]; $p = .002$).

Having added these tests, we then evaluated the effect of removing tests to minimize the number of tests and optimize the predictive power. Removing Trails A and B resulted in a model that was significantly better at classifying at-risk and safe drivers than the original Clinical Model (Trails A removed: IDI [95% CI] = .21 [.07 - .34], $p = .003$; Trails B removed: IDI [95% CI] = .22 [.08 - .36], $p = .002$).

Thus, our final Improved Model included contrast sensitivity, UFOV subtest 2, visual acuity and MMSE (but not Trails A and B; Table 5). It had significantly better predictive power than the original Clinical Model (AUC .91; IDI [95% CI] = .21 [.07 - .34], $p = .003$; Figure 3). The main gain was in sensitivity, which improved from .50 to .80 at the optimal cut point, while specificity was very high for both models (Table 5).

4. DISCUSSION

4.1. MOT and MoCA

Both MOT and MoCA were clearly predictive of driving performance: safe drivers had higher (better) MOT thresholds and higher MoCA scores than at-risk drivers. However, neither test provided enough *new* information to warrant inclusion in our clinical test battery at this time. In general, the brief MOT underperformed as a predictor, relative to the other measures we tested. Specificity was high (safe drivers were very likely to have high tracking thresholds) but, even with a Youden-optimal cutoff point, at-risk drivers were just as likely to "fail" the brief MOT as to "pass" it. Brief MOT scores were strongly correlated with the UFOV subtest 2 (Table 3), which is unsurprising, since these tests share a divided attention component. However, brief MOT was less predictive of driving scores than the UFOV subtest 2, indicating that the test was not bringing additional driving-relevant information to bear. We suggest that this failure reflects only our specific implementation of MOT. We chose to measure threshold speed, which might not be as important as how long one could sustain tracking, or how many objects could be tracked simultaneously.

The MoCA fared better than the brief MOT. Unsurprisingly, there was a strong correlation between the MoCA and the MMSE, since they measure similar cognitive skills. Indeed, the correlation between the MoCA and the MMSE was the highest pairwise correlation in the study ($\rho = 0.60$; Table 3). Although the ROC curves in Figure 2b suggest a trend for the MoCA to be a better predictor than the MMSE, the MoCA did not yield a significantly greater AUC than the MMSE (Table 4). For the difference in the AUCs to be significant, at least 80 participants would have been required (power 80% and $\alpha = 5\%$). Based on the results of this pilot study we can only conclude that the MoCA appears to be equivalent to the MMSE in terms

of success at predicting driving performance and there is no evidence that one is superior to the other in our data. Clearly, however, a larger sample study is required before firm conclusions can be drawn.

4.2. Optimizing the DriveWise clinical assessment battery

The baseline Clinical Model, comprising tests from the standard DriveWise clinical assessment battery, performed reasonably well, with an AUC of .79. However, it did not substantially outperform the MMSE by itself; nor did substituting the MoCA for the MMSE improve matters. Adding the brief MOT also did not change model performance. This may be due to the fact that the Clinical Model included Trails B, which is often employed to test functions such as scanning and divided attention (Lezak *et al.* 2004, Greenstein *et al.* 2009). The brief MOT may be redundant with Trails B. The brief MOT was more strongly correlated with Trails B than with any other measure in our battery, save the UFOV subtest 2 (Table 3).

Given the strong performance by the UFOV subtest 2 by itself, it is not surprising that adding this subtest to the battery improved performance substantially over the baseline Clinical Model. The final Improved Model (which included visual acuity, contrast sensitivity, MMSE, and UFOV subtest 2) yielded an impressive AUC of .91. Our interpretation is that the UFOV subtest 2 is adding important information which is not redundant with the MMSE and the vision tests. As a result of this study, DriveWise have now added UFOV to their clinical assessment battery.

4.3. Comparison to other studies

How does our Improved Model compare to other published batteries? We identified a set of recent studies that report ROC data for predicting road test performance using more than one test (Table 1). Two useful benchmarks are Bédard *et al.*'s study (2008), who used three of the key

tests we examined (MMSE, Trails A, and UFOV), and a recent retrospective study by O'Connor et al. (2010). This latter study tested the predictive ability of an interview-based screening tool, the "4Cs", which asks patients about crash history, family concerns, clinical status, and cognitive problems. While the 4Cs tool does not include an attention test, it is relevant because it was applied to the same population as our study - participants in the DriveWise program (O'Connor et al. 2008).

Our results agreed quite well with Bédard et al.'s (2008). In their sample, the AUCs for the MMSE, Trails A, and UFOV subtest 2 were .71, .65, and .82, respectively, while ours were .68, .64, and .84. This suggests that our sampling population performed similarly to Bédard et al.'s (2008), which was also a convenience sample.

The O'Connor et al. (2010) study reported 160 participants with a mean age of 80.5 years. The 4Cs yielded an AUC of .81 for classifying subjects into safe (pass) vs. at-risk (fail/marginal) on the road-test outcome, compared to the .91 AUC of our Improved Model. While the ability to correctly predict at-risk drivers was lower for the 4Cs than our Improved Model (sensitivity .58 vs. .80, respectively), the ability to correctly predict safe drivers was similar (specificity .94 vs. .95).

Unfortunately, beyond the two studies just described, reporting the AUC does not appear to be standard practice, even when other ROC values are reported. Instead the practice is to report the optimal sensitivity and specificity. While these values are useful, they allow for some ambiguity. Consider Table 5. If we compare the Improved Model to the MoCA Model, the Improved Model has a greater specificity *and* sensitivity, so the Improved Model dominates the MoCA model. However, if we compare the basic Clinical Model to the MoCA Model (which swaps the MMSE for the MoCA), the situation is less clear. The Clinical Model has a very high

specificity (1.0) but a low sensitivity (.50), while the MoCA Model has a lower specificity (0.91) and a higher sensitivity (0.60). Which model is better? Unless we have an explicit cost function to assign values to classifying an unsafe driver as safe and classifying a safe driver as unsafe, sensitivity and specificity are insufficient. We need the AUC measure, which tells us that the Clinical Model is superior (.79 versus .73).

Where only sensitivity and specificity were reported, we estimated the AUC by computing A, the area under the one-point ROC (Zhang and Mueller 2005). This measure, a refinement of measures such as A' and A'', has problems (Macmillan and Creelman 1996), but is the best we can do for a one-point comparison to the AUC. For our data, A tends to overestimate the underlying AUC (Table 6). We compared A values for our study (Table 6) and recent comparable studies (Table 1) (Hoffman *et al.* 2005, Oswanski *et al.* 2007, Wood *et al.* 2008, Kay *et al.* 2009, Korner-Bitensky and Sofer 2009, Dobbs and Schopflocher 2010, Carr *et al.* 2011, Wood *et al.* 2013). One takeaway is that a wide range of tests can achieve a reasonable A score. Unsurprisingly, batteries perform better than individual tests. The DriveAble screen (Korner-Bitensky and Sofer 2009), Wood *et al.*'s battery (2008, 2013), and the DriveSafe/DriveAware combination (Kay *et al.* 2009) all achieve A scores in the high .80s, comparable to our baseline Clinical Model. However, our Improved Model, with a score of .93, performs a little better. Since this is not a formal meta-analysis, we cannot determine whether this improvement would be deemed statistically significant. Nevertheless, it is a promising outcome. Furthermore, differences in the characteristics of the populations tested (e.g., clinical, non-clinical, or mixed) also need to be considered as these may affect estimates of sensitivity and specificity.

4.4 Study limitations

This was a pilot study that used a convenience sample primarily comprising older drivers with mild cognitive impairments referred to a clinical driving assessment program. As such, the results can neither be generalized to non-clinical populations of older drivers, nor to populations of older drivers with other types of impairments (such as motor limitations or visual impairments). Furthermore, our sample size was limited and some participants who arrived late did not have time to perform all of the screening tests. Only one of the evaluators in the road test was masked to all of the scores on the screening battery. However, the second evaluator, although not completely masked, was unaware of the scores for those tests that were the main focus of the study. Despite these limitations, which are typical of a pilot study, our preliminary findings for the Improved Model are promising and worthy of further investigation in a larger-sample study.

There is an implicit assumption in this type of study that a road test is the “gold standard” and is sufficient for determining safe and unsafe drivers. However, participants’ performance might not be representative of their habitual driving: due to nervousness and an unfamiliar vehicle, their driving could be worse than usual, or their driving might be better than usual, as they know that they are being tested.

5. CONCLUSIONS

In this preliminary investigation we introduced two tests not previously evaluated for predicting at-risk drivers. Although the MoCA performed equivalently to the widely-used MMSE, a follow-up study with a larger sample is needed to confirm our findings. The brief MOT did not add any new information not captured by existing attention tasks. Additionally, we systematically evaluated a set of batteries composed from the tests at our disposal. The Improved

Model outperformed the other combinations we tested in predicting the at-risk outcome and would take less than 30 minutes to administer in a clinical driving assessment program.

However, there is still room for improvement. The Improved Model, despite an AUC of .91 and an A of .93, accounts for only 60% of the variance in road test performance. Most importantly, while the Improved Model is good at identifying safe drivers (high specificity), with false positive rates of only 5%, it still fails to identify 20% of at-risk drivers. This single data point illustrates how much work there is still to do in this field.

One avenue for future research would be to evaluate whether the sensitivity of our Improved Model could be increased with the addition of tests to address other aspects of driving, such as awareness of traffic situations and one's own driving, as evaluated by the DriveAware/Drive Safe combination, which has already demonstrated high sensitivity in clinical populations including drivers with mild cognitive impairment (Kay *et al.* 2009).

FUNDING

This work was supported in part by the National Institutes of Health (grant numbers R00 EY018680 and #1 UL1 RR 025758-02 [a pilot grant from Harvard Catalyst, The Harvard Clinical and Translational Science Center]).

ACKNOWLEDGMENTS

The authors would like to thank Mark Whitehouse for assessing driving performance, and Lissa Kapust and DriveWise personnel for logistical support in conducting the study.

CONFLICTS OF INTEREST

None of the authors have any conflicts of interest.

REFERENCES

- Anstey, K.J., Wood, J., Lord, S., Walker, J.G., 2005. Cognitive, Sensory and Physical Factors Enabling Driving Safety in Older Adults. *Clinical Psychology Review* 25 (1), 45-65.
- Arditi, A., 2005. Improving the Design of the Letter Contrast Sensitivity Test. *Investigative Ophthalmology & Visual Science* 46 (6), 2225-2229.
- Asimakopulos, J., Boychuck, Z., Sondergaard, D., Poulin, V., Menard, I., Korner-Bitensky, N., 2012. Assessing Executive Function in Relation to Fitness to Drive: A Review of Tools and Their Ability to Predict Safe Driving. *Australian Occupational Therapy Journal* 59 (6), 402-427.
- Ball, K.K., Beard, B.L., Roenker, D.L., Miller, R.L., Griggs, D.S., 1988. Age and Visual Search: Expanding the Useful Field of View. *Journal of the Optical Society of America A, Optics and image science* 5 (12), 2210-9.
- Ball, K.K., Roenker, D.L., Wadley, V.G., Edwards, J.D., Roth, D.L., Mcgwin, G., Raleigh, R., Joyce, J.J., Cissell, G.M., Dube, T., 2006. Can High-Risk Older Drivers Be Identified through Performance-Based Measures in a Department of Motor Vehicles Setting? *Journal of the American Geriatrics Society* 54 (1), 77-84.
- Bédard, M., Weaver, B., Darzins, P., Porter, M.M., 2008. Predicting Driving Performance in Older Adults: We Are Not There Yet! *Traffic injury prevention* 9 (4), 336-41.
- Bowers, A.R., Anastasio, J., Howe, P.D.L., O'connor, M.G., Hollis, A., Kapust, L., Bronstad, P.M., Horowitz, T.S.H., 2011. Dynamic Attention as a Predictor of Driving Performance in Clinical Populations: Preliminary Results. *Sixth International Driving Symposium on Human Factors in Driving Assessment, Training, and Vehicle Design. Lake Tahoe, CA.* 307-313
- Bronstad, P.M., Bowers, A.R., Albu, A., Goldstein, R.B., Peli, E., 2013. Driving with Central Field Loss I: Effect of Central Scotomas on Responses to Hazards. *JAMA Ophthalmology* 131 (3), 303-309.
- Carr, D.B., Barco, P.P., Wallendorf, M.J., Snellgrove, C.A., Ott, B.R., 2011. Predicting Road Test Performance in Drivers with Dementia. *Journal of the American Geriatrics Society* 59 (11), 2112-2117.
- Clay, O.J., Wadley, V.G., Edwards, J.D., Roth, D.L., Roenker, D.L., Ball, K.K., 2005. Cumulative Meta-Analysis of the Relationship between Useful Field of View and

- Driving Performance in Older Adults: Current and Future Implications. *Optometry & Vision Science* 82 (8), 724-731.
- Crizzle, A.M., Classen, S., Bedard, M., Lanford, D., Winter, S., 2012. MMSE as a Predictor of on-Road Driving Performance in Community Dwelling Older Drivers. *Accident Analysis and Prevention* 49, 287-292.
- Dobbs, B.M., Schopflocher, D., 2010. The Introduction of a New Screening Tool for the Identification of Cognitively Impaired Medically at-Risk Drivers: The SIMARD a Modification of the Demtect. *Journal of Primary Care & Community Health* 1 (2), 119 - 127.
- Dougherty, B.E., Flom, R.E., Bullimore, M.A., 2005. An Evaluation of the Mars Letter Contrast Sensitivity Test. *Optometry and Vision Science* 82 (11), 970-975.
- Edwards, J., Ross, L., Wadley, V., Clay, O., Crowe, M., Roenker, D., Ball, K., 2006. The Useful Field of View Test: Normative Data for Older Adults. *Archives of Clinical Neuropsychology* 21 (4), 275-286.
- Edwards, J.D., Lunsman, M., Perkins, M., Rebok, G.W., Roth, D.L., 2009. Driving Cessation and Health Trajectories in Older Adults. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences* 64 (12), 1290-1295.
- Edwards, J.D., Vance, D.E., Wadley, V.G., Cissell, G.M., Roenker, D., Ball, K.K., 2005. Reliability and Validity of Useful Field of View Test Scores as Administered by Personal Computer. *Journal of Clinical and Experimental Neuropsychology* 27 (5), 529-543.
- Federal Highway Administration Department of Transportation (Us), 1999. Highway statistics 1999. Washington (DC): FHWA. Available at: <http://www.fhwa.dot.gov/ohim/hs99/tables/dl20.pdf>.
- Federal Highway Administration Department of Transportation (Us), 2009. Highway statistics 2009. Washington (DC): FHWA. Available at: <http://www.fhwa.dot.gov/policyinformation/statistics/2009/dl22.cfm>.
- Fluss, R., Faraggi, D., Reiser, B., 2005. Estimation of the Youden Index and Its Associated Cutoff Point. *Biometrical Journal* 47 (4), 458-472.
- Folstein, M.F., Folstein, S.E., Mchugh, P.R., 1975. "Mini-Mental State:" A Practical Method for Grading the Cognitive State of Patients for the Clinician. *Journal of Psychiatric Research* 12 (3), 189-198.

- Fonda, S.J., Wallace, R.B., Herzog, A.R., 2001. Changes in Driving Patterns and Worsening Depressive Symptoms among Older Adults. *Journals of Gerontology Series B- Psychological Sciences and Social Sciences* 56 (6), S343-S351.
- Freitas, S., Simoes, M.R., Alves, L., Santana, I., 2013. Montreal Cognitive Assessment Validation Study for Mild Cognitive Impairment and Alzheimer Disease. *Alzheimer Disease & Associated Disorders* 27 (1), 37-43.
- Friedman, C., Mcgwin, G., Ball, K.K., Owsley, C., 2013. Association between Higher Order Visual Processing Abilities and a History of Motor Vehicle Collision Involvement by Drivers Ages 70 and Over. *Investigative Ophthalmology & Visual Science* 54 (1), 778-782.
- Greenstein, Y., Blachstein, H., Vakil, E., 2009. Interrelations between Attention and Verbal Memory as Affected by Developmental Age. *Child Neuropsychology* 16 (1), 42-59.
- Hoffman, L., Mcdowd, J.M., Atchley, P., Dubinsky, R., 2005. The Role of Visual Attention in Predicting Driving Impairment in Older Adults. *Psychology and Aging* 20 (4), 610-622.
- Hunt, L.A., Murphy, C.F., Carr, D., Duchek, J.M., Buckles, V., Morris, J.C., 1997. Reliability of the Washington University Road Test - a Performance-Based Assessment for Drivers with Dementia of the Alzheimer Type. *Archives of Neurology* 54 (6), 707-712.
- Kay, L.G., Bundy, A.C., Clemson, L., Cheal, B., Glendenning, T., 2012. Contribution of Off-Road Tests to Predicting on-Road Performance: A Critical Review of Tests. *Australian Occupational Therapy Journal* 59 (1), 89-97.
- Kay, L.G., Bundy, A.C., Clemson, L.M., 2009. Predicting Fitness to Drive in People with Cognitive Impairments by Using Drivesafe and Driveaware. *Archives of Physical Medicine and Rehabilitation* 90 (9), 1514-1522.
- King-Smith, P.E., Grigsby, S.S., Vingrys, A.J., Benes, S.C., Supowit, A., 1994. Efficient and Unbiased Modifications of the Quest Threshold Method: Theory, Simulations, Experimental Evaluation and Practical Implementation. *Vision Research* 34 (7), 885-912.
- Korner-Bitensky, N., Bitensky, J., Sofer, S., Man-Son-Hing, M., Gelinas, I., 2006. Driving Evaluation Practices of Clinicians Working in the United States and Canada. *American Journal of Occupational Therapy* 60 (4), 428-434.

- Korner-Bitensky, N., Sofer, S., 2009. The Driveable Competence Screen as a Predictor of on-Road Driving in a Clinical Sample. *Australian Occupational Therapy Journal* 56 (3), 200-205.
- Kunar, M.A., Carter, R., Cohen, M., Horowitz, T.S., 2008. Telephone Conversation Impairs Sustained Visual Attention Via a Central Bottleneck. *Psychon Bull Rev* 15 (6), 1135-40.
- Kundu, S., Aulchenko, Y.S., Van Duijn, C.M., Janssens, A., 2011. Predictabel: An R Package for the Assessment of Risk Prediction Models. *European Journal of Epidemiology* 26 (4), 261-264.
- Lezak, M.D., Howieson, D.B., Loring, D.W., Hannay, H.J., Fischer, J.S., 2004. *Neuropsychological Assessment*, 4th ed. Oxford University Press, New York, NY.
- Lyman, S., Ferguson, S. A., Braver, E. R. and Williams, A. F., 2002. Older driver involvements in police reported crashes and fatal crashes: trends and projections. *Injury Prevention* 8 (2), 116-120.
- Macmillan, N.A., Creelman, C.D., 1996. Triangles in Roc Space: History and Theory of "Nonparametric" Measures of Sensitivity and Response Bias *Psychonomic Bulletin & Review* 3 (2), 164-170.
- Marottoli, R.A., Mendes De Leon, C.F., Glass, T.A., Williams, C.S., Cooney, L.M., Berkman, L.F., Tinetti, M.E., 1997. Driving Cessation and Increased Depressive Symptoms: Prospective Evidence from the New Haven Epese. *Journal of the American Geriatrics Society* 45 (2), 202-206.
- McGwin, J., Gerald and Brown, D. B., 1999. Characteristics of traffic crashes among young, middle-aged, and older drivers. *Accident Analysis and Prevention* 31 (3), 181-198.
- Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., Chertkow, H., 2005. The Montreal Cognitive Assessment, Moca: A Brief Screening Tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society* 53 (4), 695-699.
- Nazem, S., Siderowf, A.D., Duda, J.E., Have, T.T., Colcher, A., Horn, S.S., Moberg, P.J., Wilkinson, J.R., Hurtig, H.I., Stern, M.B., Weintraub, D., 2009. Montreal Cognitive Assessment Performance in Patients with Parkinson's Disease with "Normal" Global Cognition According to Mini-Mental State Examination Score. *Journal of the American Geriatrics Society* 57 (2), 304-308.

- O'Connor, M.G., Kapust, L., Hollis, A., 2008. Drivewise: An Interdisciplinary Hospital-Based Driving Assessment Program. *Gerontology & Geriatrics Education* 29 (4), 351-362.
- O'Connor, M.G., Kapust, L.R., Lin, B., Hollis, A.M., Jones, R.N., 2010. The 4cs (Crash History, Family Concerns, Clinical Condition, and Cognitive Functions): A Screening Tool for the Evaluation of the at-Risk Driver. *Journal of the American Geriatrics Society* 58 (6), 1104-1108.
- Oswanski, M.F., Sharma, O.P., Raj, S.S., Vassar, L.A., Woods, K.L., Sargent, W.M., Pitock, R.J., 2007. Evaluation of Two Assessment Tools in Predicting Driving Ability of Senior Drivers. *American Journal of Physical Medicine and Rehabilitation* 86 (3), 190-199.
- Owsley, C., Ball, K., Mcgwin, G., Sloane, M.E., Roenker, D.L., White, M.F., Overley, E.T., 1998. Visual Processing Impairment and Risk of Motor Vehicle Crash among Older Adults. *Journal of the American Medical Association* 279 (14), 1083-1088.
- Owsley, C., Mcgwin Jr, G., 2010. Vision and Driving. *Vision Research* 50 (23), 2348-2361.
- Pencina, M.J., D'agostino, R.B., D'agostino, R.B., Vasan, R.S., 2008. Evaluating the Added Predictive Ability of a New Marker: From Area under the Roc Curve to Reclassification and Beyond. *Statistics in medicine* 27 (2), 157-172.
- Petersen, R.C., 2004. Mild Cognitive Impairment as a Diagnostic Entity. *Journal of Internal Medicine* 256 (3), 183-194.
- Pylyshyn, Z.W., Storm, R.W., 1988. Tracking Multiple Independent Targets: Evidence for a Parallel Tracking Mechanism. *Spatial Vision* 3 (3), 179-97.
- R Development Core Team, 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reitan, R.M., 1955. The Relation of the Trail Making Test to Organic Brain Damage. *Journal of Consulting Psychology* 19 (5), 393-394.
- Risacher, S.L., Wudunn, D., Pepin, S.M., Magee, T.R., McDonald, B.C., Flashman, L.A., Wishart, H.A., Pixley, H.S., Rabin, L.A., Pare, N., Englert, J.J., Schwartz, E., Curtain, J.R., West, J.D., O'Neill, D.P., Santulli, R.B., Newman, R.W., Saykin, A.J., 2013. Visual Contrast Sensitivity in Alzheimer's Disease, Mild Cognitive Impairment, and Older Adults with Cognitive Complaints. *Neurobiology of Aging* 34 (4), 1133-1144.

- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., 2011. Proc: An Open-Source Package for R and S+ to Analyze and Compare Roc Curves. *BMC Bioinformatics* 12 (1), 77.
- Van Calster, B., Van Huffel, S., 2010. Integrated Discrimination Improvement and Probability-Sensitive Auc Variants. *Statistics in medicine* 29 (2), 318-319.
- Wood, J.M., Anstey, K.J., Kerr, G.K., Lacherez, P.F., Lord, S., 2008. A Multidomain Approach for Predicting Older Driver Safety under in-Traffic Road Conditions. *Journal of the American Geriatrics Society* 56 (6), 986-993.
- Wood, J.M., Anstey, K.J., Lacherez, P.F., Kerr, G.K., Mallon, K., Lord, S.R., 2009. The on-Road Difficulties of Older Drivers and Their Relationship with Self-Reported Motor Vehicle Crashes. *Journal of the American Geriatrics Society* 57 (11), 2062-2069.
- Wood, J.M., Horswill, M.S., Lacherez, P.F., Anstey, K.J., 2013. Evaluation of Screening Tests for Predicting Older Driver Performance and Safety Assessed by an on-Road Test. *Accident Analysis and Prevention* 50, 1161-1168.
- Zhang, J., Mueller, S., 2005. A Note on Roc Analysis and Non-Parametric Estimate of Sensitivity. *Psychometrika* 70 (1), 203-212.

Table 1. Summary of recent comparable studies predicting road test performance of older drivers with A values, the area under the one-point ROC, to provide a summary measure for overall comparison.

Study	Comments	N	Mean Age (y)	Population	% failed road test	Specificity	Sensitivity	A
O'Connor et al., (2010)	4Cs interview-based screening tool	160	80.5	Driver evaluation program 58% progressive neurological conditions (early Alzheimer's and Parkinson's)	69	.94	.58	.85
Oswanski, et al., (2007)	Motor-free visual perceptual test (MVPT)	232	≥ 55 ^a	Driver evaluation program	44	.60	.83	.79 ^b
	Clock Drawing Task			Referrals based on unsafe behaviors observed by law enforcement officers				
Korner-Bitensky et al., (2009)	DriveABLE screen battery combines six subtests	52	71.6	Driver evaluation program 58% cognitive decline & neurological conditions	81	.90	.76	.89
Wood et al., (2008)	Final battery included vision tests, cognitive tests, and motor performance (e.g., knee extension strength)	270	75.8	Non-clinical Recruited from general population of older drivers	17	.70	.91	.88

Wood et al., (2013)	Multidisciplinary screening battery (same as (Wood <i>et al.</i> 2008))	79	72.2	Non-clinical Recruited from general population of older drivers	34	.73	.80	.83
	Screening battery plus hazard perception test			(from electoral roll)		.78	.85	.87
Kay et al., (2009)	Two tests: DriveSafe, a visual task reporting details about traffic on a rotary; and DriveAware, a driving awareness questionnaire	96	62.2 ^c	Driver evaluation program 60% Dementia, MCI and neurological conditions	38 ^d	.58	.97	.87 ^d
Hoffman et al., (2005) ^e	DriverScan			Non-clinical		.65	.71	.74
	UFOV			Recruited from general population of older drivers	N/A	.52	.85	.77
	DriverScan UFOV	155	75.2	(database of study participants)		.46	.88	.77
	DriverScan&UFOV					.71	.68	.75
Carr et al., (2011)	Test combination of Snellgrove Maze Task, Eight-Item Informant Interview to Differentiate Aging and Dementia, and Clock Drawing Task	99	74.2	Driver evaluation program 100% diagnosis of dementia	65	.80	.77	.85 ^f
Dobbs and Schopflocher (2010)	SIMARD (Screen for the Identification of cognitively impaired Medically At-Risk Drivers) comprises 3 tasks from the DemTect test for detecting mild cognitive impairment and dementia	425	76.2	Driver evaluation program and general population of older drivers 80% cognitive impairment (with or without dementia)	50	.91	.45	.79 ^g

^a Mean age is not reported; all drivers were ≥ 55 years

^b Figure 1 of the Oswanski et al. paper lists the AUC for the MVPT as .758, with a 95% confidence interval from .698 to .811. No AUC is listed for the Clock Drawing Task, however.

^c This number is not quite correct: Kay et al. started off with 115 participants, with a mean age of 62.2 y. However, 19 of these are excluded from the driving assessment task because they were either learner drivers or had primarily physical deficits. They do not report the mean age for either the excluded subset or the included 96 subjects.

^d Kay et al. did not report their results in a format compatible with our analysis. They reported specificity and sensitivity separately for two groups (test and validation), and for two different cutoffs (low and high, corresponding to a trichotomization strategy). We averaged the scores for the two groups (weighted by N), and report the greatest A value (which turns out to be for the low cutoff score), and the fail rate for the low cutoff score.

^e This study did not use a road test, but accident rates and simulator performance. Neither measure was able to predict accidents better than chance. Also, Hoffman et al. reported sensitivity and specificity only for the "most liberal cutpoint" yielding "best sensitivity" for each item, rather than optimizing the sensitivity-specificity tradeoff. If the ROC curves were symmetrical, this would not matter. It is difficult to tell from their Figure 3 whether or not this is the case, since the aspect ratio of the figure is substantially greater than 1:1.

^f Did not report sensitivity and specificity, so the values were estimated from their figure of the ROC curve for the model

^g Similar to Kay et al, Dobbs and Schopflocher reported data separately for two groups (initial and validation), and for two different cutoffs (low and high, corresponding to a trichotomization strategy). We averaged the scores for the two groups (weighted by N), and report the greatest A value (which turns out to be for the low cutoff score).

Table 2. Characteristics and Test Battery Results for Safe and At-Risk Participants

	Safe		At-risk		p-value
	<i>n</i> ^c	Mean (SD) or %	<i>n</i> ^d	Mean (SD) or %	
Demographics					
Age (years)	29	79.4 (8.7)	18	81.8 (8.4)	.57
Sex, male	29	59%	18	39%	.19 ^b
Mild cognitive impairment	29	10%	18	44%	.008 ^b
Vision tests					
Visual acuity (logMAR) ↓	29	0.10 (0.12)	18	0.11 (0.09)	.89
Contrast sensitivity (log units) ↑	28	1.61 (0.13)	15	1.55 (0.10)	.10
Cognitive tests					
MMSE (points) ↑	29	28.14 (1.55)	18	26.22 (3.02)	.03
MoCA (points) ↑	25	25.60 (2.18)	15	21.20 (5.67)	.003
Trail making tasks (s) ↓					
Part A	28	38.82 (9.13)	18	45.94 (14.24)	.11
Part B	27	103.85 (48.33)	16	147.00 (62.24)	.02
Attention tests					
MOT (°/s) ↑	29	9.54 (3.83)	18	7.53 (4.72)	.23
UFOV tasks (ms) ↓					
Subtest 1	29	23.83 (13.18)	14	58.50 (76.13)	.06
Subtest 2	29	139.62 (113.84)	14	332.36 (148.81)	< .001

^a *p*-values based on the Wilcoxon rank-sum test unless otherwise noted.

^b *p*-values based on χ^2 test.

^c Number of participants out of 29 who completed the test.

^d Number of participants out of 18 who completed the test.

↑ Higher scores indicate better performance.

↓ Higher scores indicate worse performance.

Table 3. Spearman correlations among age, cognitive and attention tests

	Age	Visual Acuity ↓	Contrast Sensitivity ↑	MMSE ↑	MoCA ↑	Trail Making Part A ↓	Trail Making Part B ↓	MOT ↑	UFOV subtest 1 ↓	UFOV subtest 2 ↓
Age	-									
Visual Acuity ↓	.20	-								
Contrast Sensitivity ↑	-.47*	-.41**	-							
MMSE ↑	-.01	-.30*	-.04	-						
MoCA ↑	-.32*	-.29	.38*	.60**	-					
Trail Making Part A ↓	.24	.22	-.14	-.38**	-.45**	-				
Trail Making Part B ↓	.17	.12	-.03	-.49**	-.59**	.48**	-			
MOT ↑	-.05	-.10	-.02	.22	.30	-.20	-.40**	-		
UFOV subtest 1 ↓	.22	.26	-.32*	-.32*	-.46**	.38*	.58**	-.36*	-	
UFOV subtest 2 ↓	-.08	.19	.00	-.48**	-.50**	.22	.53**	-.50**	.48**	-

* $p < .05$ ** $p < .01$

↑ Higher scores indicate better performance.

↓ Higher scores indicate worse performance.

Table 4. Results of ROC analyses for demographic, vision, cognitive and attention measures

	Best Threshold	Specificity	Sensitivity	AUC	95% CI	Sample Size	
						Safe ^a	At-Risk ^b
Age (years)	87.5	0.86	0.33	0.55	0.38-0.72	29	18
Mild Cognitive Impairment	NA	0.90	0.44	0.67	0.54-0.80	29	18
Vision Tests							
Visual Acuity (logMAR) ↓	0.03	0.41	0.78	0.52	0.36-0.68	29	18
Contrast Sensitivity (log units) ↑	1.65	0.43	0.93	0.65	0.49-0.81	29	15
Cognitive Tests							
MMSE (points) ↑	27.50	0.76	0.67	0.68	0.51-0.86	29	18
MoCA (points) ↑	24.50	0.68	0.80	0.78	0.61-0.95	25	15
Trail Making Tasks (s) ↓							
Part A	52.00	0.93	0.39	0.64	0.48-0.81	28	18
Part B	111.50	0.67	0.75	0.72	0.56-0.87	27	16
Attention Tests							
MOT (°/s) ↑	8.07	0.86	0.44	0.60	0.43-0.78	29	18
UFOV Tasks (ms) ↓							
Subtest 1	25.00	0.79	0.50	0.66	0.49-0.83	29	14
Subtest 2	191.50	0.72	0.93	0.84	0.72-0.97	29	14

^a Number of participants out of 29 who completed the test.

^b Number of participants out of 18 who completed the test.

↑ Higher scores indicate better performance.

↓ Higher scores indicate worse performance.

Table 5. Logistic Regression Models Predicting At-Risk and Safe Drivers

	Coefficients				R ²	Hosmer-Lemeshow			AUC	95% CI	Best Threshold	
	Estimate	SE	z-value	p-value		χ^2	df	p-value			Specificity	Sensitivity
Clinical Model					0.40	7.09	8	0.53	0.79	0.61 - 0.97	0.50	1.00
Intercept	-16.38	10.82	-1.51	0.13								
Visual Acuity	0.08	0.08	1.00	0.32								
MMSE	0.61	0.35	1.75	0.08								
Trail Making Task Part A	-0.02	0.06	-0.42	0.67								
Trail Making Task Part B	0.00	0.01	-0.30	0.76								
MoCA Model					0.31	7.29	8	0.51	0.73	0.50 - 0.95	0.60	0.91
Intercept	-3.89	6.31	-0.62	0.54								
Visual Acuity	0.05	0.07	0.76	0.44								
MoCA	0.20	0.20	1.02	0.31								
Trail Making Task Part A	-0.01	0.05	-0.23	0.82								
Trail Making Task Part B	-0.01	0.01	-0.73	0.47								
MOT Model					0.40	7.06	8	0.53	0.79	0.61 - 0.97	0.79	0.82
Intercept	-17.49	11.46	-1.53	0.13								
Visual Acuity	0.08	0.08	1.01	0.31								
MMSE	0.63	0.36	1.75	0.08								
Trail Making Task Part A	0.05	0.14	0.32	0.75								
Trail Making Task Part B	-0.03	0.06	-0.47	0.64								
MOT	0.00	0.01	-0.20	0.85								
Improved Model					0.60	11.34	8	0.18	0.91	0.77-1.04	0.80	0.95
Intercept	-26.48	15.81	-1.67	0.09								
Visual Acuity	0.14	0.09	1.48	0.14								
Contrast Sensitivity	7.93	5.09	1.56	0.12								
MMSE	0.50	0.38	1.29	0.20								
UFOV Subtest 2	-0.01	0.01	-1.93	0.05								

Note. There were 10 "at-risk" participants and 22 "safe" participants included in all the logistic regression models

Table 6. A values (the area under the one-point ROC) for vision, cognitive and attention measures as well as test batteries (“models”)

Test	Specificity	Sensitivity	AUC	A ^a
Vision Tests				
Visual Acuity	.41	.78	.52	.66
Contrast Sensitivity	.43	.93	.65	.80
Cognitive Tests				
MMSE	.76	.67	.68	.78
MoCA	.68	.80	.78	.81
Trail Making Tasks				
Part A	.93	.39	.64	.79
Part B	.67	.75	.72	.77
Attention Tests				
MOT	.86	.44	.60	.75
UFOV Tasks				
Subtest 1	.79	.50	.66	.72
Subtest 2	.72	.93	.84	.89
Batteries				
Clinical Model	1.00	.50	.79	.88
MoCA Model	.91	.60	.73	.84
MOT Model	.82	.79	.79	.86
Improved Model	.95	.80	.91	.93

^a A values were calculated using the optimal specificity and sensitivity values from Table 4 (individual tests) and Table 5 (models).

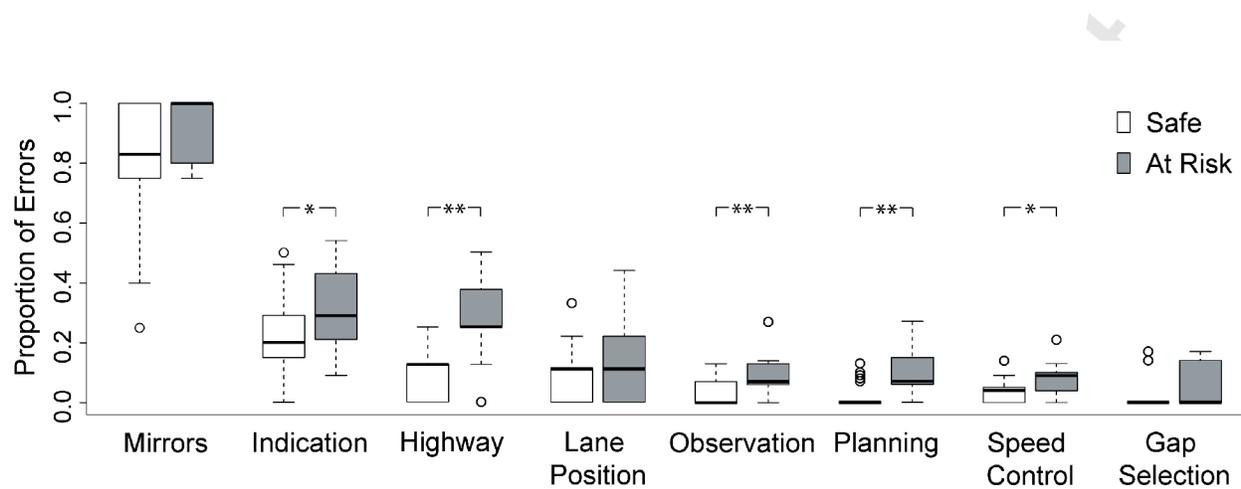
Figure Captions

Figure 1. Proportion of errors in each driving behavior category for participants considered at-risk or safe

Figure 2. Receiver operating characteristic (ROC) curves for: A, Age and vision measures; B, Cognitive tests; and C, Visual attention measures.

Figure 3. Receiver operating characteristic (ROC) curves for the original clinical model and the improved model.

Figure 1



* $p < .05$ Wilcoxon rank-sum test.

** $p < .01$

Figure 2

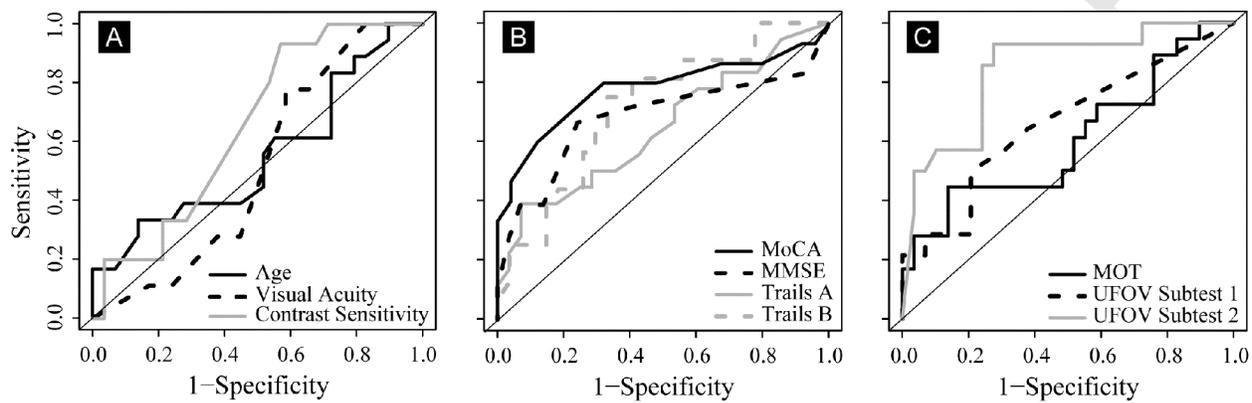
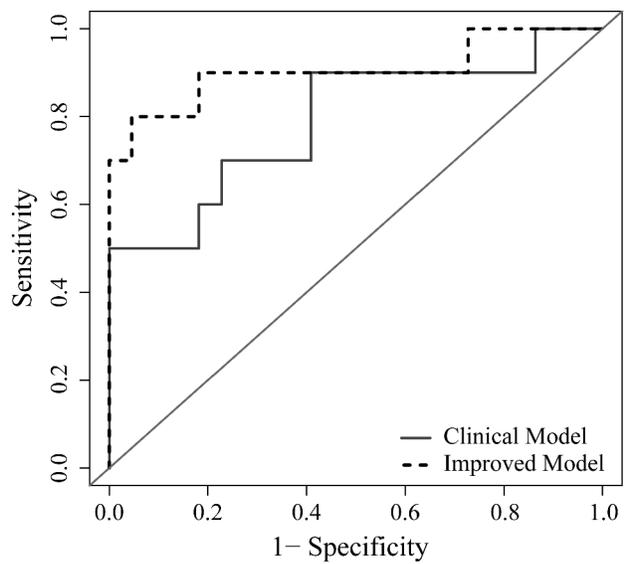


Figure 3



Appendix

Table A.1 Test Battery Results for Participants with Normal Cognition and Mild Cognitive Impairment for the Two Recruitment Groups

	Schepens		DriveWise				<i>p</i> -value
	Normal		Normal		Cognitive Impairment		
	<i>n</i>	Mean (SD)	<i>n</i>	Mean (SD)	<i>n</i>	Mean (SD)	
Vision Tests							
Visual Acuity (logMAR) [↓]	15	0.06 (0.02)	21	0.12 (0.03)	11	0.14 (0.03)	0.15
Contrast Sensitivity (log units) [↑]	15	1.53 (0.43)	20	1.54 (0.14)	8	1.60 (0.05)	0.23
Cognitive Tests							
MMSE (points) [↑]	15	28.40 (1.06)	21	28.57 (1.08)	11	23.82 (1.99)	0.00
MoCA (points) [↑]	15	25.93 (1.67)	17	25.18 (2.74)	8	17.63 (5.10)	0.00
Trail Making Tasks (s) [↓]							
Part A	14	39.00 (11.05)	21	40.76 (9.35)	11	46.55 (15.92)	0.38
Part B	14	99.43 (42.71)	20	108.70 (41.00)	9	176.67 (74.58)	0.03
Attention Tests							
MOT (°/s) [↑]	15	10.58 (3.96)	21	8.79 (3.57)	11	6.28 (4.93)	0.08
UFOV Tasks (ms) [↓]							
Subtest 1	15	21.13 (9.55)	19	31.32 (32.35)	9	66.44 (86.17)	0.08
Subtest 2	15	105.60 (103.64)	19	200.42 (120.72)	9	367.78 (158.91)	0.00

Note. All analyses performed with the Kruskal-Wallis test.

[↑] Higher scores indicate better performance.

[↓] Higher scores indicate worse performance.

Table A.2. Results of receiver operating characteristic (ROC) analyses for demographic, vision, cognitive and attention measures with the same number of participants for each test

	Best Threshold	Specificity	Sensitivity	AUC	95% CI
Age (years)	90.50	1.00	0.20	0.55	0.32 - 0.78
Mild Cognitive Impairment	NA	0.95	0.40	0.68	0.52 - 0.84
Vision Tests					
Visual Acuity (logMAR) ↓	0.12	0.41	0.80	0.55	0.35 - 0.74
Contrast Sensitivity (log units) ↑	1.65	0.45	0.90	0.58	0.38 - 0.77
Cognitive Tests					
MMSE (points) ↑	27.50	0.77	0.80	0.77	0.57 - 0.97
MoCA (points) ↑	24.50	0.73	0.7	0.72	0.49 - 0.95
Trail Making Tasks (s) ↓					
Part A	52.00	0.91	0.40	0.65	0.44 - 0.85
Part B	120.50	0.68	0.70	0.73	0.54 - 0.92
Attention Tests					
MOT (°/s) ↑	7.94	0.86	0.40	0.60	0.38 - 0.82
UFOV Tasks (ms) ↓					
Subtest 1	25.00	0.77	0.50	0.60	0.40 - 0.80
Subtest 2	191.50	0.73	0.90	0.83	0.68 - 0.98

Note. Each measure has 22 "safe" participants and 10 "at-risk" participants; participants who did not complete all tests were excluded.

↑ Higher scores indicate better performance.

↓ Higher scores indicate worse performance.

Table A.3. Logistic regression models predicting at-risk and safe drivers with different numbers of participants in each model.

	Coefficients				R ²	Hosmer-Lemeshow			AUC	95% CI	Best Threshold		Sample Size	
	Estimate	SE	z-value	p-value		χ^2	df	p-value			Specificity	Sensitivity	Safe	At-Risk
Clinical Model					0.23	5.05	8	0.75	0.74	0.57 - 0.90	0.81	0.63	27	16
Intercept	4.94	6.57	0.75	0.45										
Visual Acuity	-0.05	0.05	-0.87	0.38										
MMSE	-0.23	0.21	-1.09	0.28										
Trail Making Task Part A	0.03	0.04	0.66	0.51										
Trail Making Task Part B	0.01	0.01	0.79	0.43										
MoCA Model					0.37	10.59	8	0.23	0.76	0.58 - 0.95	0.91	0.71	23	14
Intercept	7.73	5.90	1.31	0.19										
Visual Acuity	-0.10	0.07	-1.47	0.14										
MoCA	-0.32	0.18	-1.72	0.09										
Trail Making Task Part A	0.04	0.05	0.89	0.37										
Trail Making Task Part B	0.00	0.01	0.14	0.89										
MOT Model					0.23	3.73	8	0.88	0.73	0.56 - 0.89	0.74	0.75	27	16
Intercept	4.37	6.78	0.65	0.52										
Visual Acuity	-0.05	0.05	-0.89	0.38										
MMSE	-0.22	0.21	-1.06	0.29										
Trail Making Task Part A	0.02	0.04	0.62	0.54										
Trail Making Task Part B	0.01	0.01	0.84	0.40										

MOT	0.02	0.08	0.32	0.75									
Improved Model					0.56	10.86	8	0.21	0.89	0.75 - 1.02	0.91	0.79	28 11
Intercept	20.24	12.82	1.58	0.11									
Visual Acuity	-0.09	0.08	-1.12	0.26									
Contrast Sensitivity	-8.75	4.84	-1.81	0.07									
MMSE	-0.29	0.29	-0.98	0.33									
UFOV Subtest 2	0.02	0.01	2.28	0.02									

Highlights

Predictors of at-risk older drivers were evaluated in an on-road pilot study

Useful field of view (UFOV) subtest 2 was the best single predictor

An optimal 4-test combination had 95% specificity and 80% sensitivity

It included: visual acuity, contrast sensitivity, UFOV-2 and Mini-Mental State Exam

Accepted Manuscript