A Meta-analysis of Brief, Non-computerized Cognitive Screening Tools for Predicting Unsafe Driving Among Older Adults

Authors

Ward Vanlaar[1]
Anna McKiernan[1]
Heather McAteer[1]
Robyn Robertson[1]
Dan Mayhew[1]
Steve Brown[1]
Erin Holmes[1]
David B. Carr[2,*]

Affiliations

[1]Traffic Injury Research Foundation, Ottawa, Ontario, Canada
[2]Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

[*]Correspondence

Email: DCarr@wustl.edu

ABSTRACT
Seniors represent the fastest growing population group in Canada and the United States. Given the expected increase in elderly drivers with dementia, efficient and effective screening for cognitive impairment will become more important. In light of this need, the objective of this study was to systematically review and meta-analyze the literature for brief, non-computerized cognitive screens that were evidenced-based and could be easily adopted in a driver license renewal or fitness-to-drive setting to determine their predictive value to identify unsafe driving. Studies were considered that examined road tests, driving simulator assessment or motor vehicle crashes as primary outcomes and 9 studies identifying 10 separate tests were identified based on our inclusion criteria. A small to medium-sized, significant pooled effect of 1.94 was found, meaning that on average, when cognitive screening tools predict a driver is unsafe, there is a 94% greater chance that this driver will indeed exhibit unsafe driving behavior. These results suggest that brief, paper and pencil cognitive screens may be feasible and efficiently adopted either during routine license renewal or during fitness to drive evaluations. Further studies are warranted in regards to their acceptability, reliability and validity before widespread dissemination.

***Keywords*:** cognitive screening; meta-analysis; older drivers; elderly drivers; multilevel model.

1. Introduction

Canada's population is aging and seniors represent the fastest growing population group in Canada, as in many other countries (*1*). In 2013, there were 36.8 million licensed drivers over the age of 65 years in the US (*2*). With an increase in this aging demographic on the roads, it is important to identify the physical, cognitive, and mental changes that can occur with advanced age that have a detrimental effect on driving safely. In particular, older drivers are at an increased risk of suffering from late-life cognitive impairment and dementia (*3, 4*).

Dementia and specifically common neurodegenerative diseases such as Alzheimer's disease (AD) increasingly contribute to both issues of traffic safety and loss of mobility. The province of Ontario in Canada estimates that they will have over 100,000 drivers with dementia on the road by 2028 (*5*). In a study where older adults were administered a well-validated brief cognitive test to detect dementia (e.g. the Short Blessed Test) (*6*), nearly 20% of those over age 80 years failed the screen (*7*). Another study estimated that 4% of current drivers over 75 years have a dementing illness (*8*). These studies probably underestimate the actual number of drivers with dementia on the road, since some older adults with memory loss continue to drive even after they are reported to have stopped or even if they failed or forgot to renew their licenses.

There is no single tool that can accurately identify an unfit cognitively impaired driver with absolute certainty (*9*). Similarly, there is no consensus about which cognitive abilities should be assessed with respect to driving performance and which tests should be used to assess these abilities (*10*). Martin, Marottoli and O'Neill (*11*) argue that the available literature regarding cognitive tools measuring driver safety fails to demonstrate the benefit of driver assessment for either preserving transport mobility or reducing motor vehicle crashes. Studies that examine older adults find some correlations with cognitive tools and impaired driving outcomes, but the associations for single tests are relatively weak with odds ratios and test characteristics not at a high enough level to make a decision on driving performance (*12*).

A number of commonly used cognitive screening tools, such as Trail Making Tests and the Rey Complex Figure Task, have been associated with on-road driving ability (*13-15*). As well, Staplin, Gish and Wagner (*16*) developed and evaluated a battery of tests, which was designed to assess functionality in elderly drivers with the purpose of detecting elevated crash risk. Results showed that the Useful Field of View Test (UFOV), Trail Making B, Delayed Recall, and the Motor Free Visual Perception Test (MFVPT) were associated with increased crash risk (*17*). However, studies to date in license renewal settings have not clearly shown the efficacy of routinely screening older adults with the possible exceptions of requiring in-person license renewal over age 80 years (*18*) and vision screening (*19*).

The purpose of conducting this study was to inform the selection of cognitive screening tool(s) to be potentially adopted by the Ontario's Ministry of Transportation's (MTO) 80 and Above License Renewal Program. This program requires drivers to go through a mandatory biennial license renewal cycle once they turn 80 years of age. At the time of this study the evaluation included vision testing, a written test, an educational session and a review of the older adult's driver license record. A road test is also included if the driver appears to pose a risk to road safety based on this review. It was MTO's intent to select, pilot, and eventually validate brief cognitive screen(s) to help improve the License Renewal Program. Thus, the objective of this research was to a) identify brief, easily adopted and administered cognitive screens in the literature that have been associated with driving impairment in older adults and b) to perform a meta-analysis of the screens to determine group efficacy.

2. Materials and Methods

2.1. Inclusion criteria and selection of studies: To identify relevant studies for this meta-analysis (*20*, *21*), a systematic review (*22*) of the literature was conducted in 2012. This involved three separate researchers (AM, HM and SB) who independently identified articles that were relevant to the project. The search was conducted using the following key words, search engines, journals, proceedings, and libraries.

2.1.1. Keywords: task performance and analysis; testing, test; self-report; screening battery; screening; screen; psychomotor performance; psychomotor; psychometrics; predictors; predictor; performance; perception; older drivers; neuropsychology; neuropsychological; metacognition; mental ability; mental; memory; geriatric assessment; functional abilities; fitness to drive; fitness; executive functioning; executive; elderly drivers; driver screening; driver assessment; disorders; disorder; decision making; decision; cognitive; cognition; attention; assessment.

2.1.2. Journals: Transportation Research Record; Transportation Research; Traffic Injury Prevention; Neurology; Journal of the International Neuropsychological Society; Journal of the American Medical Association; Journal of the American Geriatrics Society; Journal of Safety Research; Journal of Occupational Therapy; Journal of Gerontology; International Journal of Geriatric Psychiatry; Injury Prevention; Human Factors; Cochrane Database for Systematic Reviews; Clinics in Geriatric Medicine, Canadian Family Physician; British Columbia Medical Journal; American Journal of Public Health; American Journal of Occupational Therapy; Alzheimer's Disease and Associated Disorders; Accident Analysis and Prevention.

2.1.3. Proceedings: Drugs and Traffic Safety; International Council on Alcohol; Transportation Research Board; Traffic and Transport Behavior Psychology; Fit to Drive; Association for the Advancement of Automotive Medicine

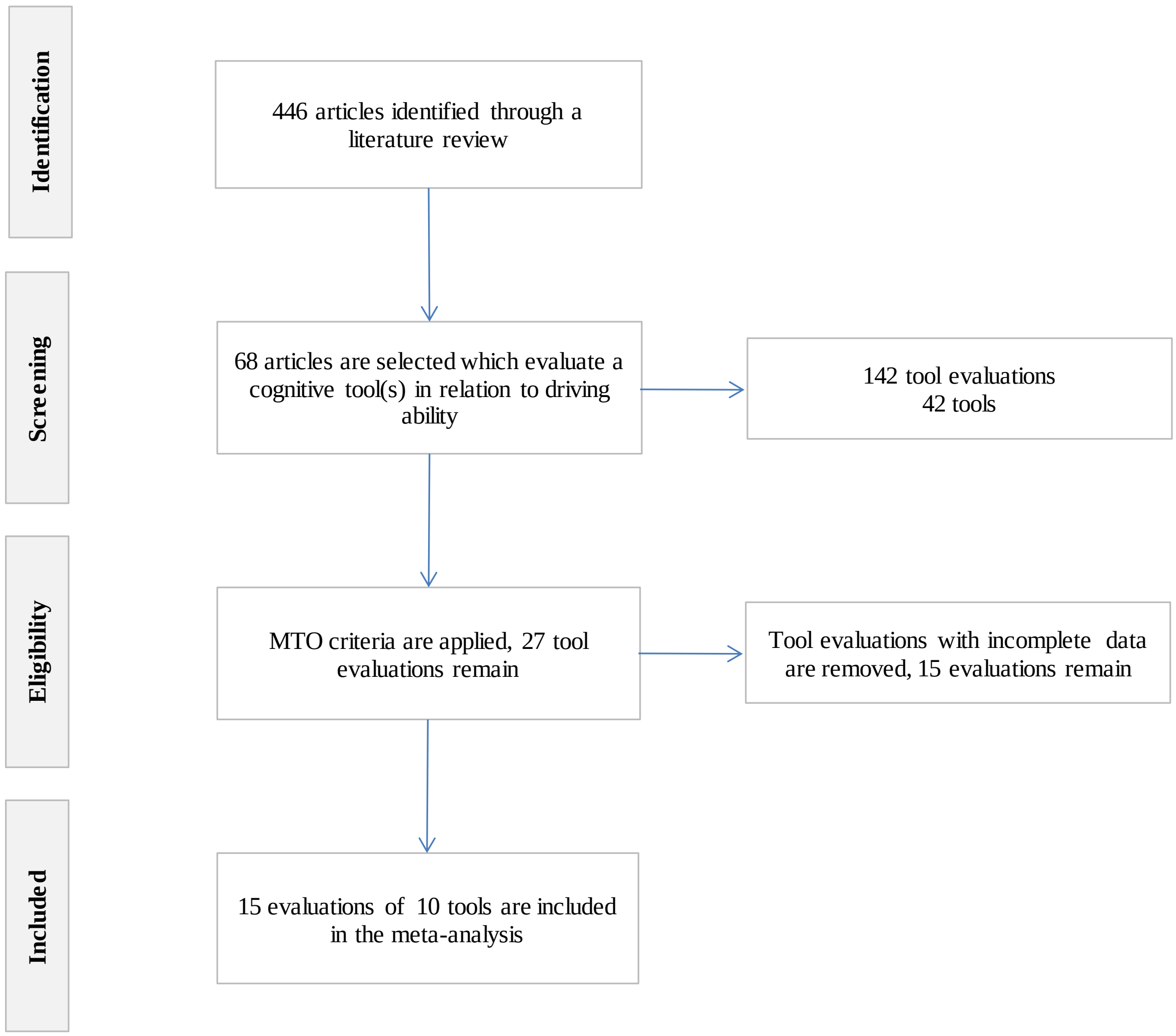
2.1.4. Search Engines and Online Catalogues: Wolters Kluwer; Springer Link; Science Direct; Scholar's Portal; Sage Journals; Safety Lit; Pubmed; PsychINFO; Proquest; Medline; Karger; Hein Online; Google; Factiva; APA PsycNet.

2.1.5. Libraries: University of Toronto; University of North Carolina Highway Safety Research Center (HSCR); University of Michigan Transportation Research Institute (UMTRI); National Highway Traffic Safety Administration (NHTSA); Insurance Institute for Highway Safety (IIHS); IMOB at University of Hasselt; SWOV Institute for Road Safety Research; Canadian Association of Road Safety Professionals (CARSP); Austroads (Australian Department of Transportation); TIRF Library.

Figure 1 contains a PRISMA diagram to visualize the selection process. The initial literature search yielded 446 articles. Articles which evaluated the predictive value of one or more cognitive screening tool(s) for older adult drivers using a performance-based road test, a driving simulator assessment, or motor vehicle crashes were extracted from the search. These articles were labeled as evaluation articles. The search identified 68 such articles which when combined contained 142 tool evaluations that were coded for the meta-analysis. Two independent researchers (AM and HM) reviewed the 142 tool evaluations and their outcomes and coded them according to a set of variables of interest to the meta-analysis. This included assigning codes in relation to MTO's needs for adopting a tool suitable for Ontario's license renewal environment. As such, several conditions had to be satisfied for a cognitive tool to be included in the meta-analysis. These inclusion criteria were based on the tools needing to meet certain characteristics:

- **Short Duration:** only screens that could be administered in less than 10 minutes were selected.
- **Group Administration:** screens were selected if appropriate for presentation to a group setting (e.g. did not require one-on-one administration)
- **Feasibility:** tools were easy to administer (i.e., simple instruction) and to score (i.e., minimal calculation needed to arrive at the score);
- **Computer:** only tools that can be administered with pencil-and-paper administration were included (e.g. computerized tests excluded).
- **Expertise:** tools that did not require a high level of specialized training for the administrators were included.
- **Cost**: although this was not an exclusion for this analysis, in general, tools that were available in the public domain would be more likely selected over those that were copyrighted and had costs that MTO would incur.

PRISMA Flow Diagram



After coding the 142 tool evaluations, researcher one (AM) and researcher two (HM) compared coding choices. If the two researchers coded the variables differently, they came to an agreement about the appropriate code through a discussion of the discrepancy and relying on a third expert (WV) when needed. Approximately 2.8% of the codes had a discrepancy that was resolved in this way. Selecting evaluations of tools that satisfied MTO's requirements using this approach with two independent coders yielded 27 tool evaluations from 15 different articles. Note that "published in a peer-reviewed journal" has not been used as an inclusion/exclusion criterion in this review. Instead, the quality of each evaluation was assessed independently by the three researchers and no critical flaws were identified in this process that would warrant the exclusion of any of the evaluations included at this stage. For instance, quality or inclusion was judged based on the participant's age group, whether a blind assessment was conducted, and if there was a control group or not.

Upon removing studies with incomplete data (for example when it was not possible to calculate a standard error, which is required for meta-analyzing the data), 15 evaluations from 9 different articles remained. These studies examined 10 different cognitive screening tools.

2.2. Cognitive screening tools included in this meta-analysis: Table 1 provides an overview of the evaluations included in this meta-analysis based on our inclusion criteria. The outcome measure used in the evaluation, along with the actual outcome and an interpretation of the result is provided, as well as a reference to the original study.

2.3. Conversion rules: From Table 1 it is apparent that the direction of the original outcome measure was not the same in every study. For instance, a higher score on the Rey-O Copy tool in one evaluation was associated with safe driving whereas a higher score on the AD-8 Information Interview in another evaluation was associated with unsafe driving. Therefore, the direction of each outcome measure was made comparable by reversing scores. Furthermore, scores were transformed into log odds ratios using rules defined by Borenstein, Hedges, Higgins and Rothstein (*20*): With studies reporting odds ratios and 95%-CIs the odds ratios were transformed into log-odds point estimates and log-odds of the 95%-CI. This was then used to calculate standard errors. For studies reporting correlation the point estimate and its 95%-CI were transformed into Cohen's d using the formula Cohen's d=(2 X correlation)/sqrt(1-(correlation X correlation)). Cohen's d results were transformed into log-odds and standard errors calculated. For studies reporting log-odds and standard errors the 95% Confidence Intervals (95%-CIs) were calculated (point estimate plus/minus 1.96 X standard error (se)). Finally, for studies reporting means data the Stata command -metan- was used to calculate Cohen's d point estimates and their 95%-CIs. These Cohen's d results were then converted into log-odds using the formula Log-odds= (Cohen's d) X (pi/sqrt(3)) and the result was used to calculate standard errors. Effect size in the meta-analysis was defined as the log-odds ratio of predicting unsafe driving behavior. Table 2 presents a summary of the converted effect sizes.

Table 1. Overview of outcome measures as reported in the original studies included in the meta-analysis.

Study	Outcome measure	Outcome	Interpretation
Goode et al. 1998	Means (s.d.; N) Crash	28.07 (7.7; 124)	Higher score on Rey-O Copy (ROCF-Copy) is associated with fewer crashes as measured by state-reported crashes
	Means (s.d.; N) Non-crash	30.57 (6.27; 115)	
Goode et al. 1998	Means (s.d; N.) Crash	8.25 (7.22; 124)	Higher score on Rey-O Immediate (ROCF-Immediate) is associated with fewer crashes as measured by state-reported crashes
	Means (s.d.; N) Fail	10.79 (7.94; 115)	
Mazer et al. 1998	Means (s.d.; N) Pass	3.9 (4.3; 33)	Failing the driving assessment is associated with more errors on the Charron Test
	Means (s.d.; N) Fail	5.4 (5.9; 51)	
Mazer et al. 1998	Means (s.d.; N) Pass	2.5 (5.1; 33)	Failing a driving assessment is associated with more errors on the single digit cancellation test (SLC)
	Means (s.d.; N) Fail	6.6 (13.0; 51)	
Mazer et al. 1998	Means (s.d.; N) Pass	5.1 (5.5; 33)	Failing a driving test is associated with more errors on the double letter cancellation test (DLC)
	Means (s.d.; N) Fail	6.2 (7.9; 51)	
MacGregor et al. 2001	Odds ratio (95%-CI)	0.88 (0.77-1.00)	Increase on Traffic Sign Recognition Test (TSRT) corresponds to a decreased crash risk as measured by state-reported crashes
Snellgrove 2005	Log odds ratio (s.e.)	-0.02 (0.01)	More time required to complete Maze Task corresponds to smaller chance of passing driving assessment
Freund et al. 2005	Correlation (95%-CI)	0.68 (0.57-0.77)	Driving errors during same-day driving assessment correlates with Clock Drawing Test scoring scale
Lafont et al. 2010	Odds ratio (95%-CI)	6.5 (2.1-20.1)	Low performance on Wechsler Digit Symbol Substitution Test (DSST) (-5.8 correct) corresponds to increased risk of unsafe driving according to composite driving indicator
Bliokas et al. 2011	Odds ratio (95%-CI)	0.882 (0.782-0.994)	Higher score on Rey-O Complex Figure Test (ROFC-Copy) corresponds to smaller chance of failing road test
Bliokas et al. 2011	Odds ratio (95%-CI)	0.964 (0.843-1.103)	Higher score on Visual Form Discrimination (VFDT) corresponds to smaller chance of failing a road test
Carr et al. 2011	Means (s.d.; N) Pass	35.2 (12.3; 35)	Failing a test is associated with requiring more than 60 seconds to complete the Maze Test - time
	Means (s.d.; N) Fail	62.5 (43.9; 65)	

Study	Outcome measure	Outcome	Interpretation
Carr et al. 2011	Means (s.d.;N) Pass	6.2 (1.2; 35)	A higher score on the Clock Drawing Test corresponds to a higher chance of passing a road test
	Means (s.d.;N) Fail	4.2 (2.5; 65)	
Carr et al. 2011	Means (s.d.; N) Pass	4.3 (1.5; 35)	Failing a driving test is associated with a higher score on the Eight Item Information Interview
	Means (s.d.; N) Fail	5.8 (1.6; 65)	
Oswanski et al. 2007	Means (s.d.; N) Capable	3.5 (0.843; 131)	A higher score on the Clock Drawing Test corresponds to a higher chance of being a capable driver
	Means (s.d.; N) Incapable	2.77 (1.159; 101)	

Table 2. Age of study populations used to evaluate the predictive value of ten cognitive screening tools.

Study	Cognitive tool	Age[1]	N[2]	log-odds[3]	se[4]
Oswanki 2007	CDT	55+	232	1.33	.25
MacGregor 2001	TSRT	65-91	120	.13	.06
Mazer 1998	Charron	61	84	.51	.41
Mazer 1998	SLC	61	84	.70	.41
Mazer 1998	DLC	61	84	.28	.41
Goode 1998	ROCF-C	55+	239	.64	.24
Goode 1998	ROCF-I	55+	239	.61	.24
Snellgrove 2005	Maze Task	65+	115	.02	.01
Lafont 2010	DSST	65-85	76	1.87	.58
Bliokas 2011	ROCF-C	61	104	.13	.06
Bliokas 2011	VFDT	61	104	.04	.07
Carr 2011	Maze Task	74.2	96	1.37	.39
Carr 2011	CDT	74.2	98	1.69	.40
Carr 2011	AD-8	74.2	99	1.74	.40
Freund 2005	CDT	60+	119	3.36	.47

1. Age ranges are indicated by two numbers separated by a hyphen or a number followed by a plus sign; average ages are indicated by one number, 2. Sample size (N), 3. Log-odds ratios, and 4. Standard errors (se)

3. Theory/Calculations

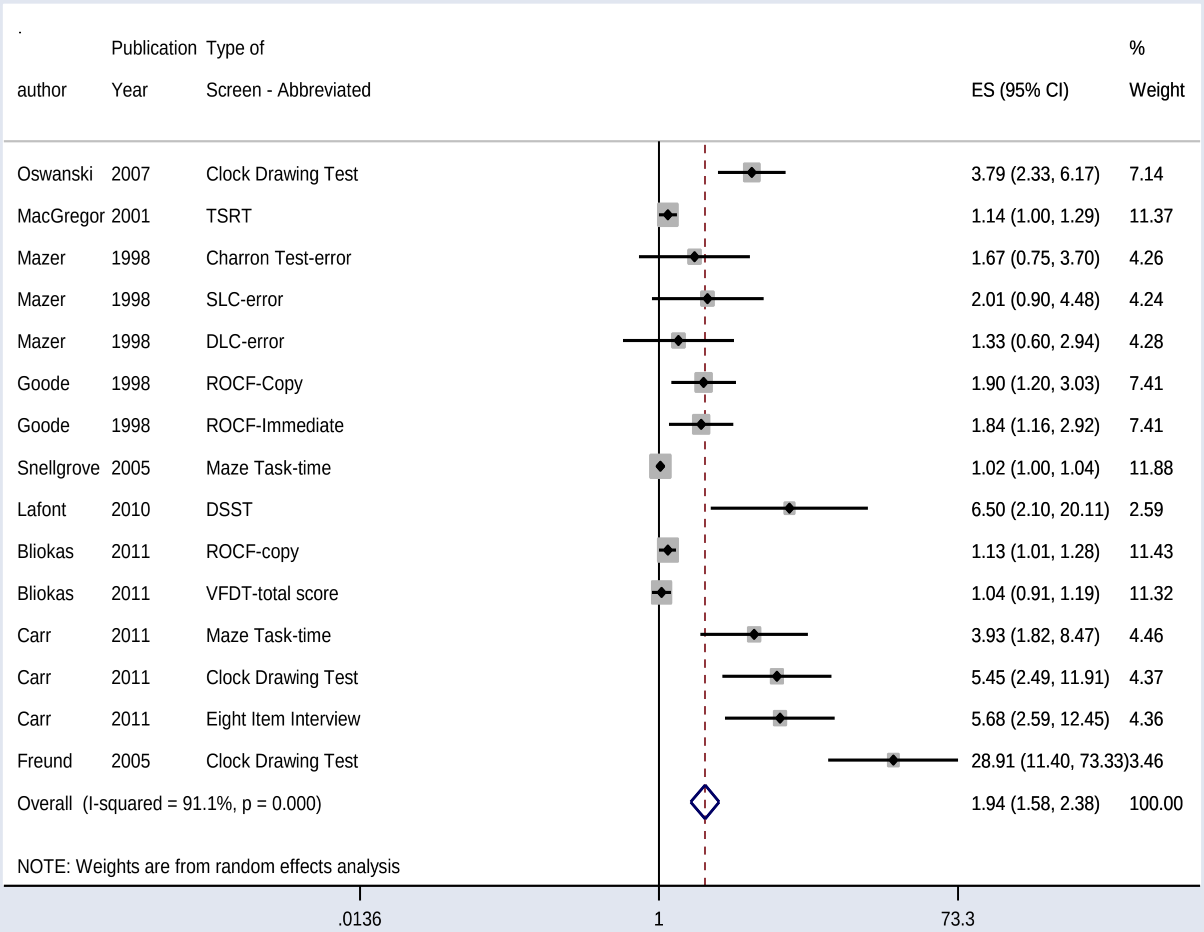
Data were analyzed in Stata, version 12 (*23*) with a random effects model using the method of DerSimonian and Laird (*24*) and the heterogeneity being estimated using the Mantel-Haenszel model (*25*). A meta-regression analysis was also conducted to explain the heterogeneity in the data, both using Residual Maximum Likelihood (REML) as well as Empirical Bayes (EB) methods (*26, 27*). Bias in the data, including publication bias, was assessed using a contour-enhanced funnel plot (*28*), Egger's test for small study effects (*25*), and Orwin's formula for calculating a Fail-Safe N (*20*). Finally, several evaluation outcomes were pooled in one study (for example, Carr, Barco, Wallendorf, Snellgrove and Ott (*29*) produced three evaluation outcomes) and such outcomes cannot be considered independent of one another. If the outcomes in these studies are positively correlated, precision may be overestimated, meaning that a significant summary effect may not be significant at all. In this case, a multilevel model can be used to properly account for such issues of dependence. Therefore, a multilevel model was estimated in MLwiN, version 2.26 (*30*) to meta-analyze the data and verify whether the results produced with the DerSimonian and Laird (*24*) model were valid (see Vanlaar (*31*) for more information about multilevel modeling).

4. Results

It should be emphasized that this meta-analysis focused on brief non-computerized tests that were readily accessible and met our specific inclusion criteria. Therefore, many tests were eliminated in the process based on our MTO criteria. Those tests that were selected were obviously cognitive based screens that typically tap into key domains such as visual scanning, search, recognition, constructional praxis, attention and/or executive function.

4.1. Overall effect size: Figure 2 is a forest plot comparing the cognitive screening tools and their ability to predict unsafe driving in older adults. The random-effects model is used as one can assume that the effect sizes truly differ across tools and studies. A significant pooled effect of 1.94 was found, meaning that on average, when cognitive screening tools predict a driver is unsafe, there is a 94% greater chance that this driver will show unsafe driving behavior (Z=6.32, p=0.000, CI 1.58-2.38). This would indicate a small to medium size effect (Cohen's d is 0.37, (*20*)). Our test for heterogeneity is significant (I-squared=91.1%, p=0.000) and 91.1% of the variation was due to heterogeneity. Thus, between-study variance is likely not only the result of random variation but rather representative of true differences in effect sizes.

Figure 2. Forest plot of the random-effects model



Variables displayed: author; publication year; type of screen – abbreviated; ES=Effect Size and their 95%-CI; percent weight of individual

4.2. Meta-regression and multilevel model: A meta-regression analysis was conducted using several covariates in an effort to help explain the heterogeneity. Results (analyses not shown) suggest that a large portion of the variation between studies can be attributed to us comparing studies that used a different methodology to study cognitive screening tools on older adult driving performance. This means there is evidence to suggest that the effect size depends on whether on-road driving assessments, simulator driving assessments, or crash records are used in the individual evaluation studies.

4.3. Bias: A contour-enhanced funnel plot was adopted to address possible publication bias. The funnel plot and a formal test for bias, Egger's test (bias coefficient=3.07, t=5.87, p=0.000, slope=-.0157), confirms there was only a small to medium bias effect.

5. Discussion

This study has identified cognitive screens that are brief, easily administered and scored, are relatively low cost, and have evidence in the literature that associates impaired performance with abnormal driving outcomes. Note that unsafe driving behavior in this study refers to unsafe performance during a road test, a simulator driving test, or state-reported crashes. The studies that are pooled in our meta-analysis used one of these three measures to assess the predictive value of a pencil-and-paper cognitive screening tool. The cognitive screens that were selected really narrowed the field of cognitive tests since studies that adopted these tools had to meet strict inclusion criteria: short duration, could be administered and scored in a group session, were feasible, non-computer based, did not require a high level of expertise, and were of low cost. These are parameters that most license renewal centers would require given limited resources in these settings.

The random-effects model in this meta-analysis study provides evidence suggesting that the cognitive screening tools identified in this systematic review adopting our specific inclusion criteria can be used to predict driving performance of older drivers. A small to medium-sized, significant pooled effect of 1.94 was found indicating a greater chance that this driver will indeed exhibit unsafe driving behavior. Thus, licensing agencies can have some confidence in adopting these tests as potential measures of unsafe driving.

Province and state license authorities struggle with the issues of screening and assessing medically impaired and especially older drivers with dementing illnesses. The utility of screening for cognitive impairment during license renewal has been questioned in the literature (e.g. Hakamies-Blomqvist, Johansson and Lundberg (*32*)) and in previous decision analysis algorithms (*33*). Performing these types of screens during fitness to drive evaluations may be more fruitful and predictive of fitness to drive.

Our objective of this study was neither to evaluate MTO's 80 and Above License Renewal Program, nor to discuss potential advantages or disadvantages of mandatory license renewal for senior drivers. This was outside the scope of this work (for an evaluation of this program, see Vanlaar, Hing, Robertson, Mayhew and Carr (*34*). These results, instead, are meant to illustrate the potential usefulness of cognitive tools as an addition to license renewal processes. These results may be helpful to licensing agencies, which are in need of an easy to use screening tool that can be cost-efficient and help inform licensing decisions. However, more studies are needed in license renewal settings and/or fitness to drive settings to validate such tests and determine the traffic safety and/or health benefit of such screening.

5.1 Limitations: A random-effects model rather than a fixed-effects model appears appropriate given there was strong evidence for heterogeneity in our study samples. It is not surprising that there is heterogeneity given the studies were varied and used different methodology. Thus, the pooled effect size is not one true effect size but an average of the different effect sizes. When using the information regarding how unsafe driving behavior was measured across individual evaluation studies (i.e., crashes, simulators, vs. on-road driving tests) as a covariate in a random-effects meta-regression analysis, this explained a majority of the between-study variance.

Further interpretation of findings regarding bias appears to coincide with expectations regarding publication of studies in this field. First, it is not surprising that there would be publication bias, as only cognitive screening tools with significant predictive value would be perceived to be of interest to the field and, as a consequence, have a higher chance of publication. Studies that would have found a result indicating that the pencil-and-paper cognitive screening tool predicts the opposite would never have been published, especially when such a result would be significant.

Nevertheless, our dataset did not exclusively consist of significant findings but did include some published results that were not significant, suggesting publication bias was likely less of a problem in our sample. Another particular concern in our sample of tool evaluations is that several of these outcomes came from the same studies. To account for this, a multilevel meta-regression model was run with consistent results suggesting that we can have confidence in the results from the meta-analysis.

A recent meta-analysis found that the best predictors of on-road driving were the Ergovision test, the UFOV, as well as tests such as paper folding and dot counting (*35*). It should be noted that many such tests may be very effectively adopted in license renewal or fitness to drive settings, but did not meet our restrictive inclusion criteria. This study did not assess the efficacy or feasibility of computer based screening tools that have some validation in the literature such as DHI (*36*) and DriveABle (*37*).

6. Conclusion

In summary our study has identified a set of cognitive tests that are inexpensive, brief, easily administered, require minimal training and could be considered as an adjunct to other screening methods in a variety of settings. Some of these tools met the needs of MTO's 80 and Above License Renewal Program and may be useful to other agencies responsible for the driver license renewal or fitness to drive testing of older drivers who are facing similar practical constraints of time and resources. Further studies and validation need to be performed focusing on important test characteristics (ROC's, Likelihood ratio's, etc.) that ultimately predict road test performance or unacceptable crash risk rates (*38*).

Acknowledgments

This research was conducted with funding from Ontario's Ministry of Transportation (MTO). The authors of this study thank Chengjie Xiong, Ph.D., Associate Professor of Biostatistics in the Department of Biostatistics, School of Medicine at Washington University for his comments on this study, which helped us to improve the quality of this work. We also thank Stephanie Paule-Martino for editing and formatting references. An abbreviated version with preliminary results of this work was presented at the 58th International Annual Meeting of the Human Factors and Ergonomics Society and published in the Proceedings (*39*).

References

1. R. Robertson, W. Vanlaar, Elderly drivers: Future challenges? *Accident Analysis & Prevention* **40**, 1982-1986 (2008).
2. Insurance Information Institute Inc. (2015).
3. B. M. Dobbs, "Medical Conditions and Driving: Current Knowledge, Final Report. Association for the Advancement of Automotive Medicine," (NHTSA Publication, 2005).
4. D. B. Carr, B. R. Ott, The older adult driver with cognitive impairment: "It's a very frustrating life". *Jama* **303**, 1632-1641 (2010).
5. R. W. Hopkins, L. Kilik, D. J. Day, C. Rows, H. Tseng, Driving and dementia in Ontario: A quantitative assessment of the problem. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie* **49**, 434-438 (2004).
6. R. Katzman *et al.*, Validation of a short Orientation-Memory-Concentration Test of cognitive impairment. *Am J Psychiatry* **140**, 734-739 (1983).
7. J. Stutts, J. Stewart, C. Martell, Cognitive test performance and crash risk in an older driver population. *Accid Anal Prev* **30**, 337-346 (1998).
8. D. J. Foley, K. H. Masaki, G. W. Ross, L. R. White, Driving cessation in older men with incident dementia. *Journal of the American Geriatrics Society* **48**, 928-930 (2000).
9. M. Bedard, B. Weaver, P. Dārzin, M. M. Porter, Predicting driving performance in older adults: we are not there yet! *Traffic Inj Prev* **9**, 336-341 (2008).
10. K. J. Anstey, J. Wood, S. Lord, J. G. Walker, Cognitive, sensory and physical factors enabling driving safety in older adults. *Clin Psychol Rev* **25**, 45-65 (2005).
11. A. J. Martin, R. Marottoli, D. O'Neill, Driving assessment for maintaining mobility and safety in drivers with dementia. *The Cochrane database of systematic reviews* **8**, (2013).
12. K. Ball *et al.*, paper presented at the Driving Assessment 2005: International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design, Rockport, Maine, June 27-30, 2005 2005.
13. R. Raedt, I. Ponjaert-Kristoffersen, The relationship between cognitive/neuropsychological factors and car driving performance in older adults. *Journal of the American Geriatrics Society* **48**, 1664-1668 (2000).
14. A. K. Schanke, K. Sundet, Comprehensive Driving Assessment: Neuropsychological Testing and On-road Evaluation of Brain Injured Patients. *Scandinavian journal of psychology* **41**, 113-121 (2000).
15. W. M. Whelihan, M. A. DiCarlo, R. H. Paul, The relationship of neuropsychological functioning to driving competence in older persons with early cognitive decline. *Archives of Clinical Neuropsychology* **20**, 217-228 (2005).
16. L. Staplin, K. W. Gish, E. K. Wagner, MaryPODS revisited: updated crash analysis and implications for screening program implementation. *J Safety Res* **34**, 389-397 (2003).
17. K. K. Ball *et al.*, Can high-risk older drivers be identified through performance-based measures in a Department of Motor Vehicles setting? *J Am Geriatr Soc* **54**, 77-84 (2006).
18. D. C. Grabowski, C. M. Campbell, M. A. Morrisey, Elderly licensure laws and motor vehicle fatalities. *Jama* **291**, 2840-2846 (2004).
19. G. McGwin, S. A. Sarrels, R. Griffin, C. Owsley, L. W. Rue, The impact of a vision screening law on older driver fatality rates. *Archives of ophthalmology* **126**, 1544-1547 (2008).
20. M. Borenstein, L. V. Hedges, J. Higgins, H. R. Rothstein, *Introduction to Meta-*

Analysis. (Wiley Online Library, West Sussex, United Kingdom, 2009).

21. A. J. Sutton *et al.*, *Methods for meta-analysis in medical research*. (J. Wiley Chichester; New York, 2000).
22. M. Egger, G. D. Smith, D. Altman, *Systematic reviews in health care: meta-analysis in context*. (John Wiley & Sons, 2008).
23. StataCorp. (StataCorp LP, College Station, TX, 2011).
24. R. DerSimonian, N. Laird, Meta-analysis in clinical trials. *Controlled clinical trials* **7**, 177-188 (1986).
25. J. Sterne, Meta-analysis in Stata: an updated collection from the Stata Journal. *Meta-Analysis In Stata: An Updated Collection From The Stata Journal-9781597180498-52, 88*, (2009).
26. R. Harbord, P. Whiting, Meta-analysis in Stata: an updated collection from the Stata Journal. (2009).
27. S. Sharp, in *Meta-Analysis In Stata: An Updated Collection From The Stata Journal*, J. A. C. Stern, Ed. (Stata Press, College Station, Texas, 2009), pp. 97-106.
28. T. M. Palmer, J. L. Peters, A. J. Sutton, S. G. Moreno, Contour-enhanced funnel plots for meta-analysis. *Stata Journal* **8**, 242 (2008).
29. D. B. Carr, P. P. Barco, M. J. Wallendorf, C. A. Snellgrove, B. R. Ott, Predicting road test performance in drivers with dementia. *J Am Geriatr Soc* **59**, 2112-2117 (2011).
30. J. Rasbash, Charlton, C., Browne, W.J., Healy, M. and Cameron, B. (Centre for Multilevel Modelling, University of Bristol, 2012).
31. W. Vanlaar, Multilevel modeling in traffic safety research: two empirical examples illustrating the consequences of ignoring hierarchies. *Traffic Inj Prev* **6**, 311-316 (2005).
32. L. Hakamies-Blomqvist, K. Johansson, C. Lundberg, Medical Screening of Older Drivers as a Traffic Safety Measure—A Comparative Finnish-Swedish Evaluation Study. *Journal of the American Geriatrics Society* **44**, 650-653 (1996).
33. S. M. Retchin, B. E. Hillner, The costs and benefits of a screening program to detect dementia in older drivers. *Med Decis Making* **14**, 315-324 (1994).
34. W. Vanlaar, M. M. Hing, R. Robertson, D. Mayhew, D. Carr, An evaluation of Ontario's Group Education Session (GES) for license renewal of seniors aged 80 and above. *J Safety Res* **56**, 33-39 (2016).
35. J. Mathias, L. Lucas, Cognitive predictors of unsafe driving in older drivers: a meta-analysis. *Internatio-nal psychogeriatrics* **21**, 637-653 (2009).
36. L. Staplin, K. W. Gish, K. H. Lococo, J. J. Joyce, K. J. Sifrit, The Maze Test: A significant predictor of older driver crash risk. *Accid Anal Prev* **50**, 483-489 (2013).
37. A. R. Dobbs, Accuracy of the DriveABLE cognitive assessment to determine cognitive fitness to drive. *Canadian family physician* **59**, e156-e161 (2013).
38. B. Weaver, S. D. Walter, M. Bedard, How to report and interpret screening test properties: Guidelines for driving researchers. *Traffic Inj Prev* **15**, 252-261 (2014).
39. W. Vanlaar, A. McKiernan, H. McAteer, R. Robertson, D. Mayhew, D. Carr, S. Brown, E. Holmes, *Cognitive screening tools for predicting unsafe driving behavior among senior drivers: a meta-analysis*. In Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting, October 27–31, Chicago, Illinois, USA. (Human Factors and Ergonomics Society, 2014, Sage Publications: Los Angeles, CA, p. 2048-2052).