HHS Public Access

The Clinical Utility of the Trail Making Tests As Predictors of Driving Performance in Older Adults

George D. Papandonatos, PhD[1], **Brian R. Ott, MD**[2], **Jennifer D. Davis, PhD**[3], **Peggy P. Barco, OTD**[4], and **David B. Carr, MD**[5]

[1]Department of Biostatistics, Brown University, Providence, RI

[2]Department of Neurology, Warren Alpert Medical School of Brown University, and Rhode Island Hospital, Providence, RI

[3]Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, and Rhode Island Hospital, Providence, RI

[4]Program in Occupational Therapy, Washington University School of Medicine in St. Louis

[5]Department of Medicine and Neurology, Washington University School of Medicine, St. Louis, MO

Abstract

OBJECTIVES—To assess the clinical utility of trail making tests as screens for impaired road-test performance.

DESIGN—We performed secondary analyses on three separate data sets from previously published studies of impaired driving in older adults using comparable road test designs and outcome measures.

SETTING—Two academic driving specialty clinics.

PARTICIPANTS—A total of 392 older drivers (303 with cognitive impairment and 89 controls) from Rhode Island and Missouri.

MEASUREMENTS—Standard operating characteristics were evaluated for Trail Making Test Part A (TMT-A), and Part B (TMT-B), as well as optimal upper and lower test cut points that could be useful in defining groups of drivers with indeterminate likelihood of impaired driving, who would most benefit from further screening or on-road testing.

Address correspondence to: George D. Papandonatos, PhD, Center for Statistical Sciences, Brown University, Box G-S121, Providence, RI 02912, Phone: 401-863-9186, FAX: 401-863-9182, gdp@stat.brown.edu.

RESULTS—Discrimination remained relatively high (>70%), when cut points for trail making tests derived from Rhode Island data were applied to Missouri data, but calibration was poor (p<.01). TMT-A provided the best utility for determining a range of scores (68–90 sec) for which additional road testing would be indicated in general practice settings. TMT-B was limited by a high frequency of cognitively-impaired participants unable to perform the test within the allotted time (>25%). Mere inability to complete the test in a reasonable time frame, e.g., TMT-A>48 sec or TMT-B>108 sec, may still be a useful tool in separating Unsafe from Safe/Marginal drivers in such samples.

CONCLUSION—Trail making tests (particularly TMT-A) may be useful as screens for driving impairment in older drivers in general practice settings, where most people are still safe drivers, but more precise screening measures need to be analyzed critically in a variety of clinical settings for testing cognitively-impaired older drivers.

INTRODUCTION

Driving fitness assessment in older adults requires functional evaluation of multiple domains (cognitive, motor, perceptual, and psychiatric). It represents a particular challenge among older adults with Alzheimer's disease, a group at increased crash risk. There have been many attempts to date to define off-road batteries employing paper and/or computerized tests that could be used to determine driving fitness or potential benefit from on-road testing. As these studies have only assessed individual samples and have tended to eschew internal cross-validation or external replication, the generalizability of their findings remains unclear.

According to the latest practice guidelines from the American Academy of Neurology on evaluating driving risk in those with dementia, "there is insufficient evidence to support or refute the benefit of neuropsychological testing, after controlling for the presence and severity of dementia."[1] While many neuropsychological tests have shown promise as screening tests in older drivers based largely on correlations with driving performance, validated cutoff scores for these tests are lacking.[2] Furthermore, there has been no single test shown to be sufficiently robust as a predictor of driving performance that it could be used as a sole screening measure to determine fitness to drive.

The Trail Making Test (TMT) is a two-part pencil-and-paper test commonly used in driving research studies due to its ease of administration and its correlation with driving errors. Part A (TMT-A)[3] is a visual-search, attention, and motor-speed task that involves connecting in numeric order a series of randomly dispersed circles containing numbers. Part B (TMT-B) additionally requires one to switch attention sets by connecting in alternating order a series of randomly dispersed circles containing numbers or letters. Although the TMT-B has been described as an executive function task, it also taps into other cognitive domains that reside outside of frontal lobe function.

TMTs have been shown to be significantly correlated with impaired driving on road tests by older drivers.[4–16] Furthermore, they have found widespread use in clinical settings as screening measures for driving impairment in older drivers with and without cognitive impairment,[17,18] having been recommended for this purpose in the past by the American Medical Association.[19] TMTs have other attractive features as potential screening tests, including their brevity (each test part taking less than five minutes), ease of administration (as pencil and paper tasks), low cost, and availability in the public domain.

Dobbs and Shergill[10] recently compared TMT to road test performance among 134 older drivers (47 cognitively impaired and 87 controls). Moderate discrimination ability was seen for time to complete both TMT-A and TMT-B. A systematic review recommended TMT-B cutoffs ranging from 90 to 180 seconds.[20] However, the tests were deemed ineffective in identifying drivers who were unfit to drive, when using binary cutoffs. This review[21], as well as another earlier review of office-based cognitive predictor tests,[22] suggested that trichotomous cutoffs should be developed to separate both safe and unsafe drivers from those in whom there is uncertainty, since it is this middle group which would benefit most from additional screening tests or on-road driving evaluation.

Given the potential utility of TMTs to screen for driving fitness in older people and the current lack of validated cutpoints, we sought to close this gap in knowledge by re-examining previous data from three separate road test studies in cognitively normal and impaired samples drawn from different locales, using similar road test designs and rating approaches. Our goals were to a) define trichotomous TMT cutoff scores that would best define marginal drivers who would benefit most from on-road testing, and b) externally validate these cutoffs by applying them to different samples of drivers.

METHODS

Participants

Cognitively impaired individuals were recruited primarily from hospital-based memory disorders centers and area physicians. Two Rhode Island patient samples (N=153), one drawn from Pawtucket and another from Providence, were evaluated at the Alzheimer's Disease & Memory Disorders Center, an outpatient diagnostic and treatment program, for participation in longitudinal studies of driving ability funded by the National Institute on Aging. Details of these studies have been reported elsewhere.[23–25] The third patient sample (N=150) was drawn from St. Louis, Missouri, evaluated in an OT-based fitness-to-drive center (the Driving Connections Clinic) and funded through the Missouri Department of Transportation, Division of Highway and Traffic Safety. Study details have been reported elsewhere.[5,26] Demographic descriptions of the samples are presented in Table 1. The samples were of similar age, education, and cognitive impairment level as measured by comparable screening measures (Short Blessed Test converted to MMSE based on a published conversion formula.[27]) All samples were obtained using similar inclusion criteria, requiring a formal diagnosis of dementia made by a physician, dementia severity in the very mild to mild range according to the Clinical Dementia Rating Scale (CDR= .5 to 1) or AD-8 (greater than or equal to 2), and an active driving license.

The Rhode Island sites also recruited N=89 cognitively healthy individuals. All were recruited from the community or were spouses of the memory-impaired participants. Healthy participants had no history of dementia and a Mini Mental State Exam score >26.

Exclusion criteria for all samples included visual acuity that did not meet state guidelines, non-English speaking, any major chronic unstable disease or condition (e.g., seizures); severe orthopedic/musculoskeletal or neuromuscular impairments that required adaptive equipment to drive; visual, hearing, or language impairments that interfered with being able to perform the testing, newly prescribed sedating drugs (e.g., use of narcotics or anxiolytics within the past month or chronic use that causes sedation); and/or a driving evaluation in the last 12 months. Rhode Island participants with a history of an at-fault accident in the prior year were excluded.

On-Road Testing

Missouri participants were administered the modified Washington University Road Test (mWURT). Rhode Island participants were administered the Rhode Island Road Test (RIRT), modeled after the original WURT. Both tests were administered by trained specialists. Although routes differed across sites, tests used comparable scoring procedures. The specialist accompanied the participant in a specially-fitted vehicle with a brake on the passenger side for emergency use, but only provided verbal instructions to complete the course. After course completion, the specialist rated the participant's driving performance as "Safe", "Marginal", or Unsafe." "Safe" implied that continued driving by the participant would be unlikely to result in crashes or violations. "Marginal" indicated that the participant could continue to drive, but should restrict driving to particular locations, times, traffic density, or enroll in driving lessons. "Unsafe" indicated that the participant exhibited driving behavior with a high probability of leading to crashes that could not be easily remediated. Studies were approved by local institutional review boards, and all participants provided signed informed consent.

Measurements: Trail Making Tests, Part A and B[3]

Trail making tests were administered according to standard procedures that included a sample trial followed by the test trial. TMT-A was discontinued at 180 seconds, and TMT-B was discontinued at 300 seconds, per test instructions. Time to complete the task in seconds was used as the primary outcome measure in data analyses. Scores were not demographically corrected.

Data Analysis

Proportional-odds logistic regression models are commonly employed to analyze ordered categorical outcomes, as their regression coefficients are readily interpretable as representing the change in the log-odds of obtaining higher, rather than lower, values of the outcome per unit change in the respective model covariate. However, power for testing the proportional-odds assumption is often limited. Therefore, we chose to employ a less-restrictive model that fits with clinical practice and allows the regression slopes to vary across thresholds for dichotomizing the outcome. This model was estimated by dichotomizing the Road Test global impression score into Unsafe vs. Marginal/Safe and

Unsafe/Marginal vs. Safe categories and fitting separate logistic regression models to each binary outcome.[28]

For ease of presentation, we only considered regression models in which the association of TMT duration with Road Test (RT) performance was evaluated separately for TMT-A and TMT-B, rather than jointly. In such single-covariate models, the sensitivity and specificity of particular TMT cutoffs remain invariant under monotone transformations of TMT duration, as they are based only on the ranks of the observations. This is also true of the Area Under the Curve (AUC) discrimination measure,[29] which estimates the probability that we can correctly rank two randomly-chosen study participants of different driving ability levels based on their TMT performance. Goodness of fit of the model was evaluated using the Hosmer-Lemeshov test.[30] Significant *p*-values are indicative of discrepancy between observed and fitted values, i.e. of poor calibration. Other typical test characteristics were calculated as well (Positive Predictive Value, Negative Predictive Value, Correct Classification Rate).

When dealing with trichotomous ordinal outcomes, it is important to clearly define what constitutes a positive vs. negative RT outcome, as these definitions underlie all subsequent test performance calculations. In our analyses, we first focused on the probability of obtaining a Safe rating, without distinguishing between Marginal and Unsafe ratings. The latter combined outcome captures all participants for whom driving concerns were identified by the instructor. We then repeated these calculations with Safe and Marginal ratings merged, and contrasted with Unsafe ratings alone. Viewed on its own, an Unsafe rating identifies participants with the most egregious driving behaviors who would be deemed unfit to drive.

In assessing TMT predictive performance, we placed particular emphasis on validation, both internal and external. For internal validation purposes, we derived RT cutoffs based on the Pawtucket sample (N=120), and sought to validate them on the Providence sample (N=122). For external validation purposes, we derived RT cutoffs based on the combined Rhode Island sample (N=242), and attempted to replicate them on the Missouri sample (N=150). External validation analyses were then repeated using cutoffs derived from cognitively impaired Rhode Island participants alone (N=153).

RESULTS

Goodness of fit of the model was improved by a logarithmic transformation of TMT duration. Tables 2–3 present TMT-A and TMT-B cutoffs chosen to ensure sensitivity around 90% for specific RT outcomes. Short TMT completion times should be associated with improved RT performance. Table 2 focuses on identifying upper limits on TMT duration met by about 90% of participants with positive RT outcomes, while Table 3 focuses on identifying lower limits on TMT duration met by about 90% of participants with negative RT outcomes. Of note, exact 90% sensitivity cutoffs could not always be attained, because of finite sample size. Also, the Rhode Island and Missouri samples showed considerable differences in driving ability, accentuating PPV differences over and above those due to between-study differences in sensitivity/specificity.

With regards to determining Safe driving, the TMT-A provided the best utility for determining a range of completion times (68–90 sec) for which additional testing would be indicated in mixed samples, with test durations shorter than 68 sec being characteristic of Safe drivers. Among cognitively impaired participants, this range shifted to 79–100 sec instead. As for Unsafe driving, test completion times between 34–48 sec for TMT-A and 83–105 sec for TMT-B would be indicative of the need for additional testing in mixed samples, with longer durations being characteristic of Unsafe drivers. At 48 sec for TMT-A and 108 sec for TMT-B, upper test cutoffs remained essentially unchanged for cognitively impaired participants alone.

The TMT-A appeared quite useful for discriminating between strong and poor road test performance in all of these settings, with the probability of correctly ranking a pair of study participants of distinct driving abilities (Unsafe vs. Not) ranging from 0.70 under internal validation in the mixed ability Rhode Island sample to 0.74 under external validation on the Missouri sample. Using patients alone as a training set degraded discrimination to 0.60 in Rhode Island, leaving Missouri results unaffected. Dichotomizing outcomes as Safe vs. Not did not materially affect these conclusions. Calibration was acceptable when evaluated on the same sample used to estimate the model. However, calibration became poor when the model was applied to independent samples, indicating the need for additional model covariates over and above TMT-A duration time.

The TMT-B had discriminatory power for the Unsafe vs. Not comparison similar to that of the TMT-A in our validation datasets, ranging from 0.70 in the mixed ability Rhode Island sample to 0.76 on the Missouri sample. Using patients alone for training purposes degraded discrimination to 0.60 in Rhode Island, leaving Missouri results unaffected. Dichotomizing outcomes as Safe vs. Not did not materially affect these conclusions. Calibration was acceptable when applying predictions from one Rhode Island sample to another, but not when using the Rhode Island samples to predict outcomes in Missouri. The defining feature of TMT-B in terms of identifying strong driving performance appeared to be whether it was completed within the allotted 300 sec time period or not, whether examining Safe drivers alone or Safe/Marginal drivers. Although cutoff sensitivity fell well below target among Rhode Island patients, it still exceeded 90% when evaluated among Missouri patients.

Exploratory analyses using combinations of TMT values (e.g., TMT-A and TMT-B minus TMT-A as joint model predictors) found no increases in predictive accuracy over use of the individual tests.

DISCUSSION

Both parts of the Trail Making Test (TMT) were evaluated in terms of their ability to predict a trichotomous measure of road test performance, as suggested by previously published recommendations.[20,22] Successive cut-offs on the high and low end of these measures using Safe vs. Marginal/Unsafe and Safe/Marginal vs. Unsafe cutoffs were of interest, as they allowed us to evaluate changes in the test operating characteristics as a result of treating marginal drivers as a separate group. Logistic regression models based on TMT performance among mixed samples of patients and healthy controls showed moderate discrimination

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

ability (AUC values >0.70), but produced predictions that were not well calibrated. The TMT-B in particular was limited by large numbers of cognitively impaired participants unable to perform the test within the allotted time, a problem that did not seem to affect TMT-A.

In order to establish a one-to-one relationship between test completion times and driving outcomes, we only presented findings from univariate logistic regression models. Multivariate models that adjusted for demographic and other participant characteristics would have improved predictive performance, at the expense of interpretability of the findings. Still, our results to date allow us to conclude that more precise screening measures need to be explored and analyzed critically across clinical settings before making policy decisions on the use of cognitive screening tests for driving impairment. Meanwhile, driving clinics and centers should validate their own specific TMT cutoffs based on the unique characteristics of their own population as well as subjective rating biases of their road test examiners.

In summary, trails tests may be useful as a screen for driving impairment in older drivers in general practice settings. When applied to drivers with dementia referred for major driving problems, the high proportion of subjects unable to complete the TMT-B within 300 sec may limit the test's utility. Still, mere inability to complete trails tests in a reasonable time frame, e.g. TMT-A>48 sec or TMT-B>108 sec, may in itself be useful in separating Unsafe from Safe/Marginal drivers in such samples, and may be suggestive of the need for further testing (e.g. performance-based road testing). Additional studies are needed to validate these findings.

ACKNOWLEDGMENTS

REFERENCES

1. Iverson DJ, Gronseth GS, Reger MA, et al. Practice parameter update: Evaluation and management of driving risk in dementia: Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2010; 74:1316–1324. [PubMed: 20385882]

2. Molnar FJ, Patel A, Marshall SC, et al. Clinical utility of office-based cognitive predictors of fitness to drive in persons with dementia: A systematic review. J Am Geriatr Soc. 2006; 54:1809–1824. [PubMed: 17198485]

3. Reitan, RM. Trail Making Test: Manual for Administration, Scoring and Interpretation. Indianapolis, IN: Indiana University Medical Center; 1958.

4. Asimakopulos J, Boychuck Z, Sondergaard D, et al. Assessing executive function in relation to fitness to drive: A review of tools and their ability to predict safe driving. Aust Occup Ther J. 2012; 59:402–427. [PubMed: 23174109]

5. Barco PP, Wallendorf MJ, Snellgrove CA, et al. Predicting road test performance in drivers with stroke. Am J Occup Ther. 2014; 68:221–229. [PubMed: 24581409]

6. Chaparro A, Wood JM, Carberry T. Effects of age and auditory and visual dual tasks on closed-road driving performance. Optom Vis Sci. 2005; 82:747–754. [PubMed: 16127341]

7. Classen S, Wang Y, Crizzle AM, et al. Predicting older driver on-road performance by means of the useful field of view and trail making test part B. Am J Occup Ther. 2013; 67:574–582. [PubMed: 23968796]

8. Dawson JD, Anderson SW, Uc EY, et al. Predictors of driving safety in early Alzheimer disease. Neurology. 2009; 72:521–527. [PubMed: 19204261]

9. De RR, Ponjaert-Kristoffersen I. Short cognitive/neuropsychological test battery for first-tier fitness-to-drive assessment of older adults. Clin Neuropsychol. 2001; 15:329–336. [PubMed: 11778771]

10. Dobbs BM, Shergill SS. How effective is the Trail Making Test (Parts A and B) in identifying cognitively impaired drivers? Age Ageing. 2013; 42:577–581. [PubMed: 23896609]

11. Grace J, Amick MM, D'Abreu A, et al. Neuropsychological deficits associated with driving performance in Parkinson's and Alzheimer's disease. J Int Neuropsychol Soc. 2005; 11:766–775. [PubMed: 16248912]

12. Hollis AM, Lee AK, Kapust LR, et al. The driving competence of 90-year-old drivers: From a hospital-based driving clinic. Traffic Inj Prev. 2013; 14:782–790. [PubMed: 24073765]

13. Koppel S, Charlton J, Langford J, et al. The relationship between older drivers' performance on the Driving Observation Schedule (eDOS) and cognitive performance. Ann Adv Automot Med. 2013; 57:67–76. [PubMed: 24406947]

14. Niewoehner PM, Henderson RR, Dalchow J, et al. Predicting road test performance in adults with cognitive or visual impairment referred to a veterans affairs medical center driving clinic. J Am Geriatr Soc. 2012; 60:2070–2074. [PubMed: 23110784]

15. Ott BR, Festa EK, Amick MM, et al. Computerized maze navigation and on-road performance by drivers with dementia. J Geriatr Psychiatry Neurol. 2008; 21:18–25. [PubMed: 18287166]

16. Ott BR, Davis JD, Papandonatos GD, et al. Assessment of driving-related skills prediction of unsafe driving in older adults in the office setting. J Am Geriatr Soc. 2013; 61:1164–1169. [PubMed: 23730836]

17. Korner-Bitensky N, Bitensky J, Sofer S, et al. Driving evaluation practices of clinicians working in the United States and Canada. Am J Occup Ther. 2006; 60:428–434. [PubMed: 16915873]

18. Dickerson AE. Driving assessment tools used by driver rehabilitation specialists: survey of use and implications for practice. Am J Occup Ther. 2013; 67:564–573. [PubMed: 23968795]

19. American Medical Association. Physician's Guide to Assessing and Counseling Older Drivers. 2010 http://www.ama-assn.org/ama/pub/category/10791.html.

20. Roy M, Molnar F. Systematic review of the evidence for Trails B cut-off scores in assessing fitness-to-drive. Can Geriatr J. 2013; 16:120–142. [PubMed: 23983828]

21. Molnar FJ, Wells GA, Mcdowell I. The derivation and validation of the Ottawa 3D and Ottawa 3DY: Three- and four-question screens for cognitive impairment. Clin Med Geriatr. 2008; 2:1–11.

22. Molnar FJ, Byszewski AM, Marshall SC, et al. In-office evaluation of medical fitness to drive: practical approaches for assessing older people. Can Fam Physician. 2005; 51:372–379. [PubMed: 15794022]

23. Davis JD, Papandonatos GD, Miller LA, et al. Road test and naturalistic driving performance in healthy and cognitively impaired older adults: Does environment matter? J Am Geriatr Soc. 2012; 60:2056–2062. [PubMed: 23110378]

24. Ott BR, Heindel WC, Papandonatos GD, et al. A longitudinal study of drivers with Alzheimer disease. Neurology. 2008; 70:1171–1178. [PubMed: 18216302]

25. Ott BR, Papandonatos GD, Davis JD, et al. Naturalistic validation of an on-road driving test of older drivers. Hum Factors. 2012; 54:663–674. [PubMed: 22908688]

26. Carr DB, Barco PP, Wallendorf MJ, et al. Predicting road test performance in drivers with dementia. J Am Geriatr Soc. 2011; 59:2112–2117. [PubMed: 22092029]

27. Meiran N, Stuss DT, Guzman DA, et al. Diagnosis of dementia. Methods for interpretation of scores of 5 neuropsychological tests. Arch Neurol. 1996; 53:1043–1054. [PubMed: 8859067]

28. Agresti, A. Categorical Data Analysis. 3 ed.. Hoboken, NJ: John Wiley & Sons Inc.; 2013.

29. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1992; 143:29–36. [PubMed: 7063747]

30. Hosmer DW, Hosmer T, Le Cessie S, et al. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med. 1997; 16:965–980. [PubMed: 9160492]

Table 1

Rhode Island and St. Louis Sample Demographics: Cognitively-Impaired Older Drivers.

Participant Characteristic	RI Pawtucket (N=78)	RI Providence (N=75)	RI Combined (N=153)	St. Louis (N=150)
Road Test; N (%)				
Safe	32 (41.0)	34 (45.3)	66 (43.1)	48 (32.0)
Marginal	35 (44.9)	32 (42.7)	67 (43.8)	15 (10.0)
Unsafe	11 (14.1)	9 (12.0)	20 (13.1)	87 (58.0)
Men; N (%)	48 (61.5)	35 (46.7)	83 (54.2)	96 (64.0)
Whites; N (%)	74 (94.9)	69 (93.2)	143 (94.1)	127 (88.8)
Age (years); mean ± SD	75.3 ± 7.1	76.6 ± 6.2	75.9 ± 0.7	73.6 ± 8.7
Education (years); mean ± SD	13.8 ± 3.4	13.5 ± 3.3	13.7 ± 3.4	15.0 ± 3.4
MMSE; mean ± SD	24.4 ± 3.4	25.1 ± 2.8	24.7 ± 3.1	25.5 ± 4.5
SBT; mean ± SD	-	-	-	8.4 ± 6.7
Trails A (sec); median (IQR)[*]	59.0 (40.5, 85.7)	59.7 (46.7, 79.6)	59.7 (44.7, 83.2)	47.5 (39.2, 77.8)
Trails A >180 sec; N (%)	3 (3.8)	1 (1.3)	4 (2.6)	3 (2.0)
Trails B (sec); median (IQR)[*]	257.5 (147.8, 301.0)	185.5 (117.4, 301.0)	209.5 (126.9, 301.0)	164.6 (108.1, 278.5)
Trails B >300 sec; N (%)	25 (37.9)	21 (28.4)	46 (32.9)	33 (25.0)

NB: Rhode Island Road Test; WURT: Washington University Road Test; MMSE: Mini-Mental State Examination; SBT: Short Blessed Test; MMSE calculations for WURT sample based upon Meiran et al.[27] SBT-to-MMSE conversion formula for subjects with possible/probable dementia; IQR: Inter-Quartile Range (1st quartile, 3rd quartile);

[*] Censored observations recoded as 181 sec (Trails A) and 301 sec (Trails B).

Table 2

Trail-Making Test Performance in Predicting Positive Road Test Outcome: Older Drivers.

Cognitive Status/ Outcome	Dataset	Rate	Test Cutoff	N	Calibration	Discrimination	Sens	Spec	PPV	NPV	CCR
All Older Drivers; Safe vs. Marginal/Unsafe	RI Pawtucket	.56	Trails A <66 sec	120	.47	.82 (.75–.90)	.91	.56	.73	.83	.76
			Trails B <300 sec	107	.32	.76 (.67–.85)	.84	.39	.67	.63	.66
	RI Providence	.61	Trails A <66 sec	122	.02	.69 (.60–.79)	.82	.40	.68	.59	.66
			Trails B <300 sec	121	.29	.73 (.64–.83)	.86	.23	.64	.52	.62
	RI Combined	.58	Trails A <68 sec	242	.79	.77 (.70–.83)	.89	.48	.70	.76	.72
			Trails B <300 sec	228	.18	.74 (.68–.81)	.85	.31	.66	.58	.64
	MO St. Louis	.32	Trails A <68 sec	150	<.01	.75 (.67–.82)	.92	.44	.44	.92	.59
			Trails B <300 sec	132	<.01	.74 (.66–.83)	.92	.34	.44	.88	.55
All Older Drivers; Safe/Marginal vs. Unsafe	RI Pawtucket	.91	Trails A <100 sec	120	.65	.83 (.71–.95)	.90	.45	.94	.31	.86
			Trails B <300 sec	107	.54	.77 (.68–.86)	.77	.29	.94	.08	.74
	RI Providence	.92	Trails A <100 sec	122	.16	.70 (.53–.87)	.95	.20	.93	.25	.88
			Trails B <300 sec	121	<.01	.70 (.54–.86)	.86	.50	.95	.24	.83
	RI Combined	.91	Trails A <90 sec	242	.16	.77 (.70–.82)	.90	.33	.93	.24	.85
			Trails B <300 sec	228	.49	.73 (.63–.84)	.81	.41	.94	.15	.78
	MO St. Louis	.42	Trails A <90 sec	150	<.01	.74 (.66–.82)	1.00	.30	.51	1.00	.59
			Trails B <300 sec	132	<.01	.76 (.68–.84)	.94	.42	.60	.88	.67
Cognitively Impaired Older Drivers; Safe vs. Marginal/Unsafe	RI Pawtucket	.41	Trails A <74 sec	78	.90	.79 (.69–.89)	.91	.54	.58	.89	.69
			Trails B <300 sec	66	.36	.64 (.50–.78)	.67	.47	.51	.63	.56
	RI Providence	.45	Trails A <74 sec	75	<.01	.58 (.45–.71)	.79	.41	.53	.71	.59
			Trails B <300 sec	74	.19	.58 (.45–.72)	.71	.27	.45	.52	.47
	RI Combined	.43	Trails A <79 sec	153	.68	.70 (.62–.78)	.89	.42	.54	.84	.63
			Trails B <300 sec	140	.15	.61 (.52–.71)	.69	.37	.48	.58	.51

Cognitive Status/ Outcome	Dataset	Rate	Test Cutoff	N	Calibration	Discrimination	Sens	Spec	PPV	NPV	CCR
	MO St. Louis	.32	Trails A <79 sec	150	<.01	.75 (.67–.82)	.96	.32	.40	.94	.53
			Trails B <300 sec	132	<.01	.74 (.66–.83)	.92	.35	.44	.88	.55
	RI Pawtucket	.86	Trails A <110 sec	78	.11	.75 (.59–.91)	.90	.36	.90	.36	.82
			Trails B <300 sec	66	.43	.61 (.47–.76)	.61	.29	.88	.08	.58
Cognitively Impaired Older Drivers	RI Providence	.88	Trails A <110 sec	75	.34	.60 (.38–.82)	.95	.22	.90	.40	.87
			Trails B <300 sec	74	<.01	.60 (.37–.82)	.75	.56	.92	.24	.73
	RI Combined	.87	Trails A <100 sec	153	.56	.69 (.55–.82)	.89	.35	.90	.33	.82
			Trails B <300 sec	140	.34	.62 (.49–.75)	.69	.43	.90	.15	.66
Safe/Marginal vs. Unsafe	MO St. Louis	.42	Trails A <100 sec	150	<.01	.74 (.66–.82)	1.00	.25	.49	1.00	.57
			Trails B <300 sec	132	<.01	.76 (.68–.84)	.94	.42	.60	.88	.67

NB: Calibration assessed via *p*-values for Hosmer-Lemeshov Goodness-of-Fit test; Discrimination assessed via Area Under the Curve (95% CI); Sens: Sensitivity; Spec: Specificity; PPV: Positive Predictive Value; NPV: Negative Predictive Value; CCR: Correct Classification Rate.

Table 3

Trail-Making Test Performance in Predicting Negative Road Test Outcome: Older Drivers.

Cognitive Status/Outcome	Dataset	Rate	Test Cutoff	N	Calibration	Discrimination	Sens	Spec	PPV	NPV	CCR
All Older Drivers	RI Pawtucket	.44	Trails A > 34 sec	120	.47	.82 (.75–.90)	.89	.48	.57	.84	.66
			Trails B > 84 sec	107	.32	.76 (.67–.85)	.91	.47	.53	.88	.64
	RI Providence	.39	Trails A > 34 sec	122	.02	.69 (.60–.79)	.90	.31	.46	.82	.54
			Trails B > 84 sec	121	.29	.73 (.64–.83)	.85	.54	.54	.85	.66
Unsafe/Marginal vs. Safe	RI Combined	.42	Trails A > 34 sec	242	.79	.77 (.70–.83)	.89	.38	.51	.83	.59
			Trails B > 83 sec	228	.18	.74 (.68–.81)	.89	.49	.53	.87	.64
	MO St. Louis	.68	Trails A > 34 sec	150	<.01	.75 (.67–.82)	.88	.35	.74	.59	.71
			Trails B > 83 sec	132	<.01	.74 (.66–.83)	.96	.21	.68	.77	.69
All Older Drivers	RI Pawtucket	.09	Trails A > 48 sec	120	.65	.83 (.71–.95)	.91	.63	.20	.99	.65
			Trails B >105 sec	107	.54	.77 (.68–.86)	1.00	.44	.11	1.00	.48
	RI Providence	.08	Trails A > 48 sec	122	.16	.70 (.53–.87)	.90	.52	.14	.98	.55
			Trails B >105 sec	121	<.01	.70 (.54–.86)	.80	.46	.12	.96	.49
Unsafe vs. Marginal/Safe	RI Combined	.09	Trails A > 48 sec	242	.16	.77 (.70–.82)	.90	.57	.17	.98	.60
			Trails B >105 sec	228	.49	.73 (.63–.84)	.88	.45	.11	.98	.48
	MO St. Louis	.58	Trails A > 48 sec	150	<.01	.74 (.66–.82)	.63	.70	.74	.58	.66
			Trails B >105 sec	132	<.01	.76 (.68–.84)	.90	.40	.62	.78	.66
Cognitively Impaired Older Drivers	RI Pawtucket	.59	Trails A > 37 sec	78	.90	.79 (.69–.89)	.89	.34	.66	.69	.67
			Trails B >108 sec	66	.36	.64 (.50–.78)	.92	.37	.63	.79	.67
	RI Providence	.55	Trails A > 37 sec	75	<.01	.58 (.45–.71)	.93	.18	.58	.67	.59
			Trails B >108 sec	74	.19	.58 (.45–.72)	.85	.23	.57	.57	.57
Unsafe/Marginal vs. Safe	RI Combined	.57	Trails A > 37 sec	153	.68	.70 (.62–.78)	.91	.26	.62	.68	.63
			Trails B >108 sec	140	.15	.61 (.52–.71)	.89	.30	.60	.70	.62

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Papandonatos et al.Page 14

Cognitive Status/ Outcome	Dataset	Rate	Test Cutoff	N	Calibration	Discrimination	Sens	Spec	PPV	NPV	CCR
	MO St. Louis	.68	Trails A > 37 sec	150	<.01	.75 (.67–.82)	.84	.42	.75	.56	.71
			Trails B >108 sec	132	<.01	.74 (.66–.83)	.86	.44	.73	.64	.70
	RI Pawtucket	.14	Trails A > 48 sec	78	.11	.75 (.59–.91)	.91	.45	.21	.97	.51
			Trails B >210 sec	66	.43	.69 (.47–.76)	.86	.46	.16	.96	.50
Cognitively Impaired Older Drivers											
	RI Providence	.12	Trails A > 48 sec	75	.34	.60 (.38–.82)	.89	.29	.15	.95	.36
			Trails B >210 sec	74	<.01	.60 (.37–.82)	.56	.60	.16	.91	.59
	RI Combined	.13	Trails A > 48 sec	153	.56	.69 (.55–.82)	.90	.37	.18	.96	.44
			Trails B >108 sec	140	.34	.62 (.49–.75)	.88	.21	.13	.93	.29
Unsafe vs. Marginal/Safe											
	MO St. Louis	.58	Trails A > 48 sec	150	<.01	.74 (.66–.82)	.63	.70	.74	.58	.66
			Trails B >108 sec	132	<.01	.76 (.68–.84)	.88	.40	.62	.76	.65

NB: Calibration assessed via *p*-values for Hosmer-Lemeshov Goodness-of-Fit test; Discrimination assessed via Area Under the Curve (95% CI); Sens: Sensitivity; Spec: Specificity; PPV: Positive Predictive Value; NPV: Negative Predictive Value; CCR: Correct Classification Rate.

J Am Geriatr Soc. Author manuscript; available in PMC 2016 November 01.