Drawing Clocks and Driving Cars

Use of Brief Tests of Cognition to Screen Driving Competency in Older Adults

Barbara Freund, PhD,[1] Stefan Gravenstein, MD,[1] Rebecca Ferris, BS,[1] Bonnie L. Burke, MS,[2] Elias Shaheen, BS[3]

[1]The Glennan Center for Geriatrics and Gerontology, and [2]Epidemiology and Biometry Core, Graduate Program in Public Health, Eastern Virginia Medical School, Norfolk, VA, USA; [3]Division of Gerontology, Medical College of Virginia, Richmond, VA, USA.

OBJECTIVE: The purpose of the study was to determine whether a new method of scoring the Clock Drawing Test (CDT) is a reliable and valid method for identifying older adults with declining driving competence.

DESIGN: Prospective cohort study.

SETTING: An outpatient driving evaluation clinic.

PARTICIPANTS: One hundred nineteen community-dwelling, active drivers with a valid driver's license, aged 60 and older referred for driving evaluation.

MAIN OUTCOME MEASURES: The CDT and a driving test using a STISIM Drive simulator.

RESULTS: The CDT showed a high level of accuracy in predicting driving simulation outcome (area under the receiver-operator curve, 0.90; 95% confidence interval, 0.82 to 0.95). CDT scoring scales were comparable and all correlations between CDT scores and driving performance were negative, implying that as the CDT score decreases, the number of errors increases. Interrater reliability of CDT scores was 0.95. Subjects scoring less than 5 out of 7 points on the CDT made significantly more driving errors, hazardous and in total ($P<.001$).

CONCLUSIONS: The CDT can help establish problems with executive function and indicate the need for a formal driving evaluation. Our CDT scoring scale is a reliable, valid, and time-effective screening tool for identifying elderly drivers in need of further evaluation.

KEY WORDS: clock drawing; driver screening; elderly drivers; fitness to drive.
DOI: 10.1111/j.1525-1497.2005.40069.x
J GEN INTERN MED 2005; 20:240–244.

Driving places demands on attention, memory, problem solving, and information processing, cognitive skills that often decline with aging. Older drivers crash at a rate second only to the youngest drivers.[1,2] Older driver involvement in fatal crashes is projected to increase 155% by 2030, accounting for 54% of the total projected increase in fatal crashes among all drivers.[3] As the population ages and the number of older drivers increases, declining driver competence becomes an urgent public health problem and a challenge for health professionals to recognize impaired driving ability in the elderly.

Declining driving competence is associated with impairments in vision, functional abilities, and cognition. However, it is especially true that cognitive impairment is overlooked in the context of a brief office visit as are issues related to driving ability. Cognitive impairment and dementia are surprisingly prevalent among older apparently healthy individuals, affecting up to one third of people aged over 65 years, yet it remains undiagnosed in 25% to 90%.[4–7] Consequently, the impact on driving ability persists unnoticed.

The responsibility for determining the driving fitness of older adults is increasingly falling upon the medical profession; however, clinicians have few tools and few data on which to base decisions about driving. A brief and time-efficient screening test is needed to assist clinicians to identify older drivers who may represent a public health hazard and need to undergo a driving evaluation.

The purpose of the study was to determine whether a new method of scoring the Clock Drawing Test (CDT) is a reliable and valid method for identifying older adults with declining driving competence.

METHODS

Subject Selection/Recruitment

One hundred nineteen community-dwelling, older active drivers with valid driver's licenses in Hampton Roads (6 cities and 9 counties in Eastern Virginia) were referred by their community physician or family member for driving evaluation. Subjects presented consecutively to the clinic over the course of 18 months. Eighteen subjects returned for multiple assessments, 16 at 6-month intervals, 1 at 5 months, and 1 at 1 month, yielding a total of 142 testing periods. All subjects gave consent for data to be used for research purposes. The Institutional Review Board for human subjects protection at Eastern Virginia Medical School approved the study.

Measures

All subjects completed 2 tests: the CDT and a driving simulation test. The CDT evaluates multiple areas of cognitive function including comprehension, memory, visuospatial abilities, abstract thinking, and executive function.[8] The CDT generally takes less than 5 minutes to administer and score. Subjects are verbally instructed to draw a clock, put all the numbers in, and set the time at 10 minutes after 11. This time is reported to be the most sensitive for detecting neurocognitive dysfunction.[9] The instruction is also written at the top of the page in 16-point font. Instructions may be repeated verbatim as needed. No cues are allowed. Self-correction is permitted. Several scoring scales have been developed for the CDT as a screen for dementia.[10–15] We chose 4 methods that are administered identically but vary in number of items and types of errors scored and do not use a predrawn circle: the Rouleau,[11] Mendez,[14] Manos,[15] and Freund. For our scale, we focused on 7 features (see Table 1) identified in preliminary data to be sig-

Accepted for publication July 1, 2004

The authors have no conflicts of interest to report. An earlier version of this article was presented at the 2002 American Geriatrics Society meeting, Washington, DC.

Address correspondence and requests for reprints to Dr. Freund: 825 Fairfax Avenue, Suite 202, Norfolk, VA 23507 (e-mail: freundbm@evms.edu).

Table 1. Freund CDT Scoring Scale

Time (3 points)	One hand points 2 (or symbol representative of 2)
	Exactly two hands
	Absence of intrusive marks, e.g., writing or hands indicating incorrect time, hand points to number 10; tic marks, time written in text (11:10; ten after eleven)
Numbers (2 points)	Numbers are inside the clock circle
	All numbers 1–12 are present, no duplicates or omissions
Spacing (2 points)	Numbers spaced equally or nearly equally from each other
	Numbers spaced equally or nearly equally from the edge of the circle

nificantly correlated with hazardous driving errors. Table 2 provides a brief comparison of the 4 CDT scoring methods. CDT test-retest reliability reported for Alzheimer's patients[16] is 0.70 to 0.78 without adjustment for cognitive ability, and no practice effect has been reported. Interrater agreement for CDTs drawn by healthy elderly and subjects with Alzheimer's disease is high (0.97) and does not differ between clinicians and nonclinicians.[14,16] For all 4 scoring methods, CDTs were assessed independently and blind to driving simulation test results by the principal investigator and an experienced research associate.

Driving performance was tested utilizing a STISIM Drive simulator (Systems Technology, Hawthorne, CA). Driving simulation, across a range of technological sophistication, appears to be a sensitive method to evaluate driving performance.[17–23] The STISIM Drive has been shown to correlate with onroad testing.[24,25] As an interactive system, it responds to driver inputs (steering, throttle, brake) and generates realistic roadway images in real time. The fixed-base driving cab has an adjustable car seat, accelerator and brake pedals, and dash with standard-size steering wheel. Three ceiling-mounted Epson (Long Beach, CA, USA) 700c projectors displayed roadway images on 3 contiguous 4×8 ft screens, providing a 135° field of view.

A 10-minute practice session allowed participants to become familiar with the simulator, and, in our experience, has been ample time for acclimation. Instructions and review of traffic rules were presented prior to the practice and test sessions. For the test situation, subjects were instructed to drive through an urban course (approximately 30 minutes), programmed to require execution of maneuvers which demonstrate the ability to drive and which emphasize conditions revealing discriminating errors (e.g., unprotected left turns, stop sign–controlled intersections). Speed limits varied, and turns, traffic lights, and traffic (vehicular and pedestrian) were presented. Performance measures included hazardous errors, traffic violations, and rule violations. Hazardous errors include

Table 2. Comparison of CDT Scoring Methods

Scoring Method	Total Score	Major Categories
Freund	7	Time setting, numbers, spacing
Manos	10	Time setting, spacing
Mendez	20	Hands, numbers
Rouleau	10	Hands, numbers, spacing

Table 3. Number of Driving Errors for Outcome Determination

	Hazardous Errors	Rule/Traffic Violations*
Safe drivers	0	0–16
Unsafe drivers	>2	5–35+

Note the quality of violations is not considered here.

crashes involving vehicles, buildings, trees, and pedestrians, running red lights, lane position errors, and turning positioning errors. Traffic violations refer to speeding 10 MPH or more above the posted speed limit and driving dangerously slowly—15 MPH or more below the speed limit. Rule violations were specific to the simulated course and included turning right on a red light, failing to turn, and turning in the wrong direction, thus providing a measure of short-term memory and the ability to learn new information.

In our clinic, subjects are judged as safe, conditional safe (restricted), or unsafe (failure) based on the number and type of driving errors committed (see Table 3). For example, to be considered safe the driver may not commit any hazardous errors. Determinations of restricted and unsafe are based on the presence and quality of hazardous errors and traffic or rule violations. While there is some overlap in number and types of errors, the determinations are based on severity of the errors, the length of time driving before any errors occurred, and the patients' ability to learn from the error (i.e., that particular error was not repeated).

Statistical Methods

We assessed interrater reliability by determining the consistency with which 2 independent raters applied our scoring scale to 129 randomly selected CDTs. Comparisons between our scoring algorithm with others, namely Rouleau, Mendez, and Manos, were made using Spearman's rank correlation. By showing the scores behave similarly with respect to the main outcome, and again with each other, we demonstrate equivalency reliability.

Criterion-related validity, demonstrating the accuracy of the clock test scoring scale, was assessed by comparing the CDT score with the number of specific types of errors during simulated driving tests using Spearman's rank correlation coefficient. Further criterion validity assessments were made using Spearman's rank correlation between our scoring scale

Table 4. Demographics

Characteristic	*N* Known	Summary
Male gender, *n* (%)	119	56 (47.1)
Median age, y (range)	119	77 (61–96)
Highest level of education, *n* (%)	65	
No high school diploma		7 (10.8)
High school diploma/GED		20 (30.8)
Some college		14 (21.5)
College degree		11 (16.9)
Some graduate school		1 (1.5)
Graduate school degree		12 (18.5)
Right-handed, *n* (%)	43	41 (95.3)
Lives alone, *n* (%)	49	23 (46.9)

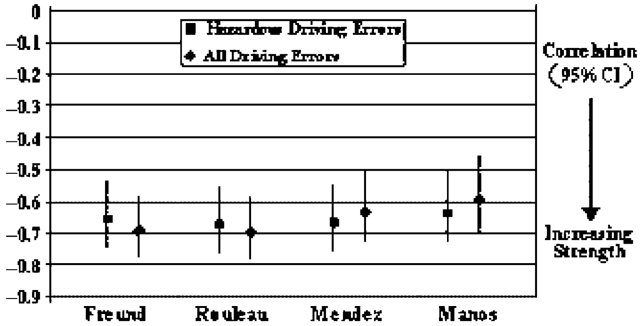

FIGURE 1. Correlation with total number of errors (hazardous and overall) by scoring scale.

and the Rouleau, Mendez, and Manos scales. Ninety-five percent confidence intervals for the rank correlation were computed based on the Fisher ρ-to-Z transformation. Differences between groups for nonparametric data were tested using Wilcoxon rank sum test with level of significance set at .05.

Receiver-operator curve (ROC) analysis was used to compare the accuracy of measures to predict driving outcome. Positive likelihood ratios (LR+) are reported with 95% confidence intervals. All analyses with the exception of the ROC analysis were conducted using SAS v8.1 (SAS Institute, Cary, NC). The ROC analysis was conducted with the aid of MedCalc v7.2.0.2 (Med Calc Software, Belgium).

RESULTS

Sample

The sample consisted of 56 (47.1%) men and 63 (52.9%) women with median age 77 years (range, 61–96 years). Table 4 provides subject demographics.

Reliability

Interrater reliability among 129 CDTs using our scoring scale was 0.95, a strong, positive correlation (95% confidence interval [CI], 0.94 to 0.97; $P<.001$). Figure 1 compares correlations for total number of hazardous driving errors and total number of all driving errors between scoring scales, with overlapping confidence intervals showing that the scales behave similarly. Directly comparing our scoring scale to other scoring scales yielded strong, positive correlations ranging between 0.75 (Freund vs Manos) and 0.82 (Freund vs Rouleau), indicating comparability of scales in general.

Validity

One hundred nine subjects completed the neuropsychological tests and had same-day driving assessments (10 subjects did not complete the driving evaluation due to simulator sickness and were excluded from this analysis). Our CDT scoring scale, modified from the other scales as described above, correlated with the number of driving errors ($r=.68$; 95% CI,0.77 to 0.57) as did the other published scoring methods (Fig. 1). Because of this comparability, further analyses with the CDT were conducted using our scoring scale.

Twenty-three (24.0%) subjects scored less than 4 out of 7 points on their CDT using our scale and all but 1 were considered unsafe by the simulated driving test. Median number of total driving errors by subjects scoring at least 4 on the CDT was 5 (interquartile range [IQR], 1–13) compared to those scoring less than 4, whose median number of total errors was 16 (IQR, 11–22.5). Median number of hazardous errors by subjects scoring at least 4 on the CDT was 0 (IQR, 0–2) compared to those scoring less than 4, whose median number of hazardous errors was 5 (IQR, 3–10.5). Note that IQR as we use it is the 25th and 75th percentiles. Subjects scoring less than 4 on our scale made significantly more driving errors, hazardous and in total ($P<.001$; Fig. 2).

A total of 138 driving assessments were performed and the overall outcome was determined to be "Safe," "Unsafe," "Restricted," or "Other/Unknown." This last category represented patients who did not complete the evaluation, were not assigned an outcome, or had missing data. In 45 (32.6%) instances the driver was considered Safe, 11 (8.0%) Restricted, 46 (33.3%) Unsafe, and 36 (26.1%) had an Other/Unknown outcome. Patients considered safe had median CDT scores of 7 (IQR, 6–7); those restricted had median CDT scores of 6 (IQR, 5–6); and patients considered unsafe had median CDT scores of 3 (IQR, 2–4).

The ability of the CDT to predict unsafe or restricted (conditional safe) driving based on driving simulation was examined. ROC analysis on the individuals with definitive driving outcomes and matching CDT evaluation showed that the CDT has a high level of accuracy to predict driving simulation outcome (area under the ROC [AUC], 0.90; 95% CI, 0.82 to 0.95), and suggested that the optimal cut point for predicting unsafe driving (Fig. 3) was a Freund score of 4, with an LR+of 27.58 (sensitivity, 64.2%; specificity, 97.7%).

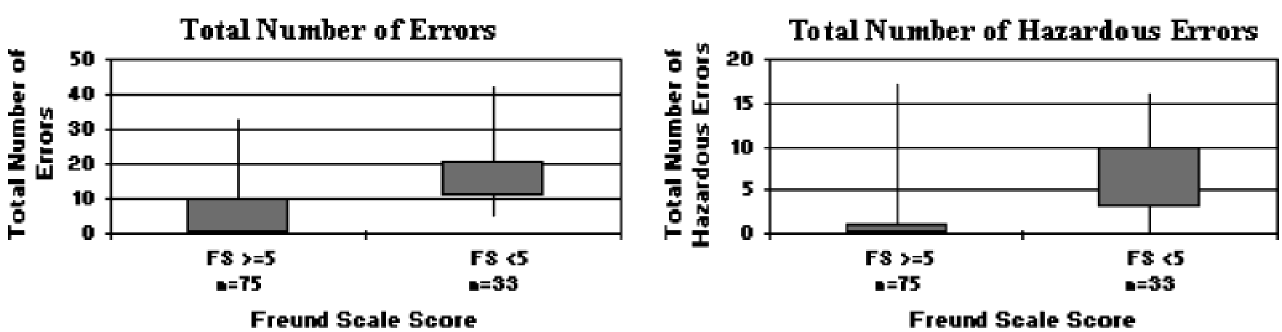

FIGURE 2. Distribution of number of driving errors by CDT (Freund scale). Box-and-whisker plots show (from top to bottom) maximum number of errors, third quartile, first quartile, and minimum number of errors. More than 2 points lost on the CDT (FS, Freund scale) was associated with significantly more errors ($P<.001$). Scores range from 0 to 7, with 7 being a perfect score.

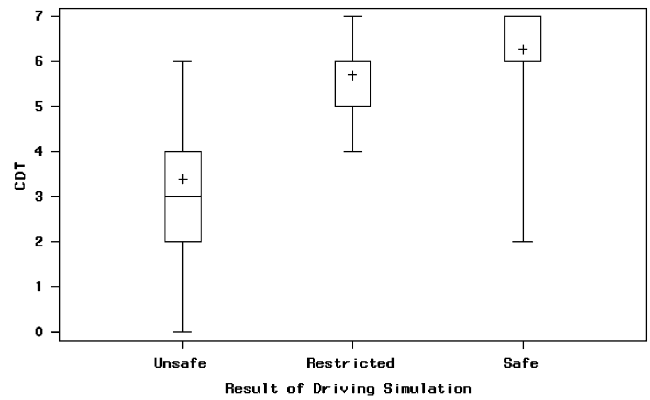

FIGURE 3. Distribution of Freund CDT by outcome of driving simulation. *Box plots of CDT score by result of driving simulation. Midline of box marks the median score, ends of the box mark the 25th and 75th percentiles, and lines extend to maximum and minimum scores. Eight of the 43 subjects who failed scored a 5, and 2 subjects scored a 6 on CDT (these 2 subjects recently immigrated to the United States and passed a subsequent evaluation 6 months later). Only 1 of 43 passers scored below a 5 (CDT score of 2), and although had no hazardous errors as defined, confused the gas and brake pedals and received 2 speeding tickets.

DISCUSSION

Driving is a complex task requiring a range of cognitive and psychomotor abilities including memory, judgment, motor control, decision making, attention, and executive function. No single short cognition test captures all these domains to a degree such that clinicians would be comfortable making absolute recommendations about a person's driving involvement. A need exists, however, for clinicians to readily distinguish between drivers in need of evaluation and those who are not.

The Clock Drawing Test is a brief, easily administered cognition test. It is commonly used in dementia evaluations but has only recently been applied to functional measures.[26,27] The CDT correlates with executive control functions, impairments of which are associated with functional disability and Alzheimer's disease.[13] The CDT also correlates with disease progression[12] and is useful in detecting very mild dementia.[15,16] Our results demonstrate that the CDT is strongly associated with driving performance. This may be because executive function is a critical component of safe driving, and in the presence of executive dysfunction, the automatized and procedural skills learned over decades of daily driving do not protect the older driver from errors. In this study, only our scoring method was analyzed to determine a cut point for driving outcomes. It should be noted that all CDT scoring methods used were strongly associated with driving errors. Thus, clinicians need not be concerned that scoring method will impact their conclusions. Our results demonstrate that the CDT is a valid screen to identify older drivers who should be referred for a driving evaluation, and it is useful as a single screen when time is limited.

Table 5. CDT Score to Predict Unsafe Driving Performance Outcome

CDT Score	Sensitivity	95% CI	Specificity	95% CI
0	1.9	0.3 to 10.1	100	91.7 to 100
1	7.5	2.1 to 18.2	100	91.7 to 100
2	20.8	10.9 to 34.1	97.7	87.7 to 99.6
3	41.5	28.1 to 55.9	97.7	87.7 to 99.6
4	64.2	49.8 to 76.9	97.7	87.7 to 99.6
5	84.9	72.4 to 93.2	76.7	61.4 to 88.2
6	96.2	87.0 to 99.4	58.1	42.1 to 73.0
7	100	93.2 to 100	0.0	0.0 to 8.3

CDT, Clock Drawing Test; CI, confidence interval.

The determination of threshold scores for CDT prediction of driver competence can assist clinicians with the issue of when to refer patients for additional testing. It is important to understand that perfect clock drawing scores do not preclude unsafe behavior, and conversely. However, scores of 4 and below on our scale are associated with unsafe driving behavior and merit additional onroad or simulator testing. Even subjects scoring 5 and 6 committed some driving errors, and, in the presence of other findings suggesting impairment, should prompt a referral for a driving evaluation.

Our findings demonstrate a moderate sensitivity (64.2%) and high specificity (97.7%) of our CDT scoring in predicting unsafe driving performance using a score of 4 as a cut point. This is significant in that an effective screening tool should not incorrectly identify someone who is able to drive competently as one who should surrender driving privileges. This is an important consideration for clinicians who practice in areas without driver evaluation resources and must decide to remove driving privileges based on clinical findings. Although the moderate sensitivity at this cut point allows for false negatives (drivers considered unsafe may score above the CDT cut point), it is highly specific for driving errors and therefore minimizes the risk for false positives (see Fig. 3). Clinicians who have driver evaluation resources available may wish to consider a higher cut point (see Table 5 for sensitivity/specificity data) when screening patients for driver evaluation referral.

Predicting driving competence has been a challenge and a burden for clinicians. Identifying unsafe drivers is a critical concern because of the implications for public safety and individual autonomy. The primary care physician is in a unique position to identify older drivers in need of driving evaluation but because of limited tools and time constraints has not fully taken on this role. The most comprehensive guide to assist clinicians in assessing older drivers, the American Medical Association's Physician's Guide to Assessing and Counseling the Older Driver, includes an earlier version of our clock scoring scale as a component of the assessment battery.[28] It appears that the CDT is an efficient instrument and a valid tool to screen for driving capability that can assist physicians in participating in this very important screening process.

Summary

Driving is an executive control task dependent on the fidelity of cognition to be performed safely. Declaring a patient unsafe to drive undermines their mobility, independence, and quality of life, and should be guided by valid measures of driving competence. The CDT is a reliable, valid, time-effective screening tool for primary care physicians to identify elderly drivers in need of further evaluation.

The authors wish to thank Myra Barnes-Eley for her assistance with the preliminary principal components analysis of the clock drawings and Robert Freund for simulator programming assistance.

This research was supported in part by The Virginia Center on Aging, Alzheimer's and Related Diseases Research Award Fund.

REFERENCES

1. **O'Neill D, Neubauer K, Boyle M, Gerrard J, Surmon D, Wilcock GK.** Dementia and driving. J R Soc Med. 1992;85:199–202.
2. **Williams AF, Carsten O.** Driver age and crash involvement. Am J Public Health. 1989;79:326–7.
3. **Lyman S, Ferguson SA, Braver ER, Williams AF.** Inj Prev. 2002;8: 116–20.
4. **Valcour VG, Masaki KH, Curb JD, Blanchette PL.** The detection of dementia in the primary care setting. Arch Intern Med. 2000;160: 2964–8.
5. **Finkel SI.** Cognitive screening in the primary care setting: the role of physicians at first point of entry. Geriatrics. 2003;58:43–4.
6. **Ross GW, Abbott RD, Petrovitch H, et al.** Frequency and characteristics of silent dementia among elderly Japanese-American men. The Honolulu-Asia Aging Study. JAMA. 1997;277:800–5.
7. **Callahan CM, Hendrie HC, Tierney WM.** Documentation and evaluation of cognitive impairment in elderly primary care patients. Ann Intern Med. 1995;122:422–9.
8. **Shulman KI.** Clock-drawing: is it the ideal cognitive screening test? Int J Geriatr Psychiatry. 2000;15:548–61; Review.
9. **Freedman M, Kaplan E, Delis D, et al.** In: Clock Drawing: A Neuropsychological Analysis. New York, NY: Oxford University Press; 1994: 3–8.
10. **Wolf-Klein GP, Silverstone FA, Levy AP, et al.** Screening for Alzheimer's disease by clock drawing. J Am Geriatr Soc. 1989;37:730–4.
11. **Rouleau I, Salmon DP, Butters N, et al.** Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. Brain Cogn. 1992;18:70–87.
12. **Sunderland T, Hill JL, Mellow AM, et al.** Clock drawing in Alzheimer's disease. A novel measure of dementia severity. J Am Geriatr Soc. 1989; 37:725–9.
13. **Royall DR, Cordes JA, Polk M.** CLOX: an executive clock drawing task. J Neurol Neurosurg Psychiatry. 1998;64:588–94.
14. **Mendez MF, Ala T, Underwood KL.** Development of scoring criteria for the clock drawing task in Alzheimer's disease. J Am Geriatr Soc. 1992;40:1095–9.
15. **Manos PJ.** Ten-point clock test sensitivity for Alzheimer's disease in patients with MMSE scores greater than 23. Int J Geriatr Psychiatry. 1999; 14:454–8.
16. **Esteban-Santillan C, Praditsuwan R, Ueda H, et al.** Clock drawing test in very mild Alzheimer's disease. J Am Geriatr Soc. 1998;46: 1266–9.
17. **Rizzo M, Reinach S, McGehee D, et al.** Simulated car crashes and crash predictors in drivers with Alzheimer disease. Arch Neurol. 1997; 54:545–51.
18. **Rizzo M, McGehee DV, Dawson JD, et al.** Simulated car crashes at intersections in drivers with Alzheimer disease. Alzheimer Dis Assoc Disord. 2001;15:10–20.
19. **Ponds RW, Brouwer WH, van Wolffelaar PC.** Age differences in divided attention in a simulated driving task. J Gerontol B Psychol Sci Soc Sci. 1988;43:151–6.
20. **Quillian WC, Cox DJ, Kovatchev BP, et al.** The effects of age and alcohol intoxication on simulated driving performance, awareness and self-restraint. Age Ageing. 1999;28:59–66.
21. **Cox DJ, Taylor P, Kovatchev B.** Driving simulation performance predicts future accidents among older drivers. J Am Geriatr Soc. 1999;47:381–2.
22. **Cox DJ, Quillian WC, Thorndike FP, et al.** Evaluating driving performance of outpatients with Alzheimer disease. J Am Board Fam Pract. 1998;11:264–71.
23. **Freund B, Risser M, Cain C, et al.** Simulated driving performance associated with mild cognitive impairment in older adults. J Am Geriatr Soc. 2001;49:S151–S152.
24. **Lee HC, Cameron D, Lee AH.** Assessing the driving performance of older adult drivers: on-road versus simulated driving. Accid Anal Prev. 2003; 35:797–803.
25. **Freund B, Gravenstein S, Ferris R, et al.** Evaluating driving competence of cognitively impaired and healthy older adults: a pilot study comparing on-road and driving simulation performance. J Am Geriatr Soc. 2002;50:1309–10.
26. **Freund B, Gravenstein S, Dobbs A, et al.** Clock Drawing Test (CDT) may predict on-road driving performance. J Am Geriatr Soc. 2001; 48:S113.
27. **Freund B, Gravenstein S, Ferris R.** Use of the Clock Drawing Test as a screen for driving competency in older adults. J Am Geriatr Soc. 2002; 50:S3.
28. **Wang CC, Kosinkski CJ, Schwartzberg JG, Shanklin AV.** Physician's Guide for Assessing and Counseling Older Drivers. Washington, DC: National Highway Traffic Safety Administration; 2003.