# Prediction of Driving Ability with Neuropsychological Tests: Demographic Adjustments Diminish Accuracy

**Joseph Barrash**,
Department of Neurology, University of Iowa

**Ashley Stillman**,
Department of Neurology, University of Iowa

**Steven W. Anderson**,
Department of Neurology, University of Iowa

**Ergun Y. Uc**,
Department of Neurology, University of Iowa and Neurology Service, Veterans Affairs Medical Center, Iowa City

**Jeffrey D. Dawson**, and
Department of Biostatistics, University of Iowa

**Matthew Rizzo**
Department of Neurology, University of Iowa

## Abstract

Demographically-adjusted norms are used to enhance accuracy of inferences based on neuropsychological assessment. However, we hypothesized that predictive accuracy regarding complex real-world activities is diminished by demographic corrections. Driving performance was assessed with a standardized on-road test in participants aged 65+ (24 healthy elderly, 26 Alzheimer's disease, 33 Parkinson's disease). Neuropsychological measures included Trailmaking A and B, Complex Figure, Benton Visual Retention, and Block Design tests. A multiple regression model with raw neuropsychological scores was significantly predictive of driving errors ($R^2 = .199$, $p < .005$); a model with demographically-adjusted scores was not ($R^2 = .113$, $p > .10$). Raw scores were more highly correlated than adjusted scores with each neuropsychological measure, and among both healthy elderly and Parkinson's patients. Demographic corrections diminished predictive accuracy for driving performance, extending findings of Silverberg and Millis (2009) that competency in complex real-world activities depends on ability levels, regardless of demographic considerations.

## Keywords

There is an established mindset in clinical neuropsychology that demographic adjustments to test scores improves the accuracy of inferences based on neuropsychological assessment (American Academy of Clinical Neuropsychology Board of Directors, 2007; Ivnik, 2004; Lezak et al., 2004). For inferences regarding the presence of acquired impairment due to brain injury or neurological disease, this emphasis on demographic adjustments is well-

Correspondence regarding this article should be addressed to Joe Barrash, PhD, Dept of Neurology, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, Iowa City, IA 52242. joseph-barrash@uiowa.edu.

supported (Lezak et al., 2004). However, sometimes the issue at hand is the adequacy of neuropsychological abilities for real-world activities. Real-world activities such as paying bills, finding one's way around town, job performance and driving safely require the same level of cognitive abilities from anyone, regardless of their demographic characteristics. Adjusted scores are informative regarding whether patients have declined from their premorbid level of ability, but for the purpose of predicting competence in real-world activities dependent on those abilities, more accurate inferences may be based on unadjusted, raw scores reflecting the individual's current level of ability. Prior to publication of large normative studies in the 1990s, studies predicting real-world functioning almost invariably employed unadjusted scores, except studies using WAIS tests. However, in their manual for demographically-adjusted norms, Heaton and colleagues (2004) recommend the use of unadjusted scores for predicting everyday functioning. When predicting competence for real-world activities, adjusting scores for demographic factors may diminish the accuracy of inferences.

This proposition has received little empirical study. In a landmark investigation, Silverberg and Millis (2009) studied 52 adults approximately one year after traumatic brain injury (TBI). Performances on seven neuropsychological tests were converted to both absolute and adjusted scores. Absolute scores were referenced to a normative sample reflecting the general adult population. Adjusted scores were based on regression models derived from healthy subjects accounting for age, education, gender and ethnicity. Overall test battery mean scores were calculated separately for absolute and adjusted scores. Outcome measures were participant or caregiver ratings for independence of living situation, community ambulation, employment, and global psychosocial functioning. Approximately one-third of the TBI participants showed non-trivial discrepancies between absolute and adjusted scores. When discrepancies were present, absolute scores were the better predictor for seven of eight outcome measures, strongly so for three of the seven: employability and two global ratings of real-world functioning. Although demonstrating the superiority of absolute scores for predicting real-world capabilities in a demographically heterogeneous sample of individuals with TBI, the authors noted limitations, and called for future research with larger and demographically distinct samples, other clinical conditions, additional outcome measures, and more reliable outcome measures.

The present study was designed to address limitations of the Silverberg and Millis (2009) study by investigating the effect of demographic adjustments on prediction of another important aspect of real-world functioning, driving. As the population of North America ages and dementia becomes more prevalent, the increasing number of elderly drivers with cognitive impairment presents a growing public health problem (Dawson et al., 2009a). However, accurate identification of unsafe drivers is important to avoid unnecessarily restricting the mobility and freedoms of competent drivers, a challenge heightened by the fact that not all patients with dementia are unsafe drivers (Hunt et al., 1993; Tallman et al., 1994).

Withaar and colleagues (2000) reviewed studies of on-road driving in older adults with cognitive decline and concluded that tests of visual scanning, visual attention and visual perception, especially the Trailmaking Test, are the most robust predictors of crashes and moving violations. Reger and colleagues (2005) conducted meta-analysis of 27 empirical studies, separately for three types of driving assessment (on-road, off-road/simulator, and caregiver report), noting that on-road tests are generally considered the "gold standard." Tests of visuospatial skills bore the strongest relationship to on-road driving performance, followed by tests of attention/concentration; tests of mental status, language and executive functions were only weakly related or unrelated. (There were insufficient studies to include memory tests in this analysis.) They concluded that tests of visual processing skills may be

most helpful for identifying at-risk drivers. Those observations have been borne out in more recent studies showing that, among patients with mild dementia, tests predicting on-road driving performance included the Trailmaking Test (Dawson et al., 2009a; Grace et al., 2005; Uc et al., 2007, 2009; Whelihan et al., 2004), the Rey Complex Figure Test copy condition (Dawson et al., 2009a; Grace et al., 2005), the Benton Visual Retention Test (Dawson et al., 2009a), and Useful Field of View (Uc et al., 2009; Whelihan et al., 2005).

In the present study, we examined the accuracy of raw versus demographically-adjusted test scores for prediction of driving ability among older drivers with very mild Alzheimer's or Parkinson's disease, and healthy elderly controls. The clinical groups were selected for very mild cognitive deficits because it is within these groups that identification of no-longer-safe drivers is especially challenging but particularly important (Tallman, 1992). To address limitations of the Silverberg and Millis (2009) study, the sample is larger, with increased clinical heterogeneity. The outcome measure is a performance-based, reliable measure of driving ability. It was hypothesized that driving performance can be predicted with raw scores from relevant neuropsychological tests, and that the accuracy of prediction is diminished when scores are adjusted for demographic factors.

## METHOD

### Participants

Participants included 24 neurologically normal elderly controls (NC), 26 individuals with probable Alzheimer's disease (AD), and 33 with Parkinson's disease (PD). All were community-dwelling, independent-living, active drivers; were age 65 or older; and had a score of 26 or higher on the mini-mental state examination (MMSE; Folstein et al., 1975). Classification as an "active driver" required a valid license and weekly driving of ten miles or more. NC were volunteers recruited from the community (primarily by advertisement at senior centers), had no neurological diagnosis or complaints, and family/collateral confirmation of the absence of abnormal cognitive decline. These were not matched groups as the study is not concerned with comparing their driving performance; rather, participants were a consecutive series of group members in the database with completed protocols and meeting criteria for the present study. AD were recruited from a registry maintained by the Department of Neurology. All were initially determined to have significant cognitive impairments and significant functional decline consistent with probable AD by clinical neuropsychological evaluation in the Benton Neuropsychology Laboratory (Tranel, 2009) prior to their identification as potential research participants, and admittance to the registry required meeting NINCDS-ADRDA criteria (McKhann et al., 1984) for diagnosis of probable Alzheimer's as determined by a board-certified neurologist. PD participants were recruited from the Movement Disorders Clinics of the University of Iowa Hospitals and Clinics and the Iowa City Veterans Affairs Medical Center, with mild-moderate disease severity by Hoehn-Yahr score. PD patients were tested during their "on" periods (the period after taking medication during which PD symptoms such as tremor, rigidity and bradykinesia are responding well, before the therapeutic effect begins waning). The sample was Caucasian other than two participants (a proportion that approximates the population of the state of Iowa): one was an African-American with AD and one a biracial individual with PD; both were highly similar to other members of their clinical group on characteristics presented below. Further details regarding these groups are available elsewhere (Dawson et al., 2009a; Uc et al., 2009).

Exclusion criteria included history of epilepsy, stroke, neoplasm, sleep disorders, depression, anxiety, schizophrenia, alcohol or substance abuse, diabetes, liver or kidney disease, severe arthritis, cancer or motion sickness, or present treatment with cognitive-dulling medications. Inclusion and exclusion criteria were carried over from prior studies

(Dawson et al., 2009a; Uc et al., 2009), with the exception of the criterion of MMSE of 26 or higher, a criterion added to provide a more stringent test of the study hypotheses (elaboration in Discussion). The study was approved by the University of Iowa Institutional Review Board, and informed consent was obtained in accord with federal guidelines for human subjects' safety and confidentiality.

## Materials and Procedure

**Neuropsychological Measures**—Five neuropsychological measures were selected for (a) their sampling of visual scanning, visual information processing and attention, areas found to correlate significantly with driving performance (Dawson et al., 2009a; Reger et al., 2004), and (b) being in widespread use in clinical practice. These included the Trailmaking Test A and B (TMT-A and -B) seconds to completion; Complex Figure Test, copy (CFT); Benton Visual Retention Test, errors (BVRT); and Wechsler Adult Intelligence Scale-III Block Design Test (BD). All tests are described in detail elsewhere (Lezak et al., 2004). Raw scores and demographically-adjusted scores were obtained for each test. Raw scores were employed rather than the absolute scores used by Silverberg and Millis (2009) because, as analyses were conducted with linear regression analysis, there was no requirement for variables to be normally distributed or on a common metric, so there was no benefit to be conferred by additional transformation of scores – yielding scores that would not necessarily be available to clinicians. Demographically-adjusted scores are expressed in *T* scores, calculated by referencing scores to published normative data adjusted for demographic factors as available from the following sources: TMT (Heaton et al., 1991), CFT (Van Gorp et al., 1990), BVRT (Benton et al., 1981), BD (Wechsler, 1997). Regarding TMT, it is noted that normative data from the Heaton et al. system are only available up to age 80, so adjustments for the 75–80 age band were applied to ten subjects over age 80.

**Driving Ability**—An instrumented vehicle known as ARGOS was employed (Rizzo et al., 1997), and driving performance was assessed with an on-road driving test, a procedure with established validity (Reger et al., 2004). Participants drove an approximately 45-minute standardized route around Iowa City in daylight, and exclusive of particularly challenging weather conditions. A trained researcher classified errors from the passenger seat. The researcher was usually blind to diagnostic group and neuropsychological test results, although group membership of PD with manifest motor symptoms would be unavoidably evident. Additionally, video of the drive was collected and reviewed a minimum of two times to assess interrater reliability (Dawson et al., 2009b). Driving ability was operationalized as the total number of errors made during the drive, with errors classified according to Iowa Department of Transportation definitions. There were 76 error types in 15 categories such as starting and pulling away from the curb, traffic signals and signs, turns, lane observations, overtaking another vehicle, control of speed, reverse driving, parking maneuvers, curves and miscellaneous events. Further details of the road test and scoring system are available elsewhere (Dawson et al., 2009b; Rizzo et al., 1997).

## RESULTS

Background characteristics of NC, AD and PD are presented in Table 1 for informational purposes only; adjustments were not made for background differences between groups because the study goal was not to compare these groups but to investigate the effect of demographic adjustments on predictive accuracy in a clinically heterogeneous sample. Study groups differed in gender composition and in age, but did not differ in years of education, MMSE, or driving errors. AD had worse scores than NC (but not PD) on four measures, and PD had worse scores than NC (but not AD) on all five measures. Driving errors were rated

with high intrarater and interrater reliability: .95 and .73, respectively (both $p < 0.001$). Driving errors were not significantly related to age in this elderly sample ($r = .18$, $p > .10$).

Preliminary analyses indicated reasonably normal distributions of independent and dependent variables, and least-squares multiple linear regression was used to test hypotheses. As the study was concerned with comparing the predictive accuracy of neuropsychological measures *in general* when demographic adjustments are or are not applied (rather than comparing the predictive value of specific measures), all five measures were entered as a block in two corresponding predictive models, one with raw scores, one with demographic adjustments. These analyses revealed that the model with raw scores from all five neuropsychological measures was significantly predictive of driving errors ($R^2 = .199$, $F = 3.68$, $p < .005$). A parallel regression analysis with demographically-adjusted scores did not account for a significant proportion of the variance in driving errors ($R^2 = .113$, $F = 1.87$, $p > .10$). A direct comparison of the $R^2$ from the two predictive models was performed with the Z statistic as described by Olkin (Kleinbaum et al., 2008) to test for the significance of the difference between correlation coefficients. This statistic, $Z = 2.09$ ($p < .05$), indicated that the model with raw scores accounted for a significantly higher proportion of variance in driving errors than did adjusted scores.

Table 2 presents the predictive relationship (Pearson correlations) of raw vs. demographically-adjusted scores from individual neuropsychological measures with driving errors. Raw scores were significantly correlated with driving errors for three of five neuropsychological measures, TMT-A, CFT and BD, and approached significance for the other two measures, TMT-B and BVRT. In contrast, only one demographically-adjusted score was significantly correlated with driving errors, TMT-A, and only two others approached significance, CFT and BD. The magnitude of the correlation was higher for raw scores than adjusted scores for all five neuropsychological measures, although the difference only reached significance for CFT and BD.

To determine whether the hypothesized effect of diminished predictive accuracy with demographic adjustments was present within each study group, separate regression analyses were conducted for each group. The $R^2$ of raw scores and demographically-adjusted scores to driving errors for each study group were .217 vs. .08 for NC, .268 vs. .280 for AD, and .161 vs. .05 for PD. In the context of small sample sizes and five predictor variables, none of these differences reached statistical significance.

## DISCUSSION

Results supported study hypotheses. Raw scores from a set of five neuropsychological measures predicted a significant 20% of the variance in driving errors in a clinically heterogeneous sample of elderly, independent-living, active drivers. Application of demographic corrections to neuropsychological scores diminished predictive accuracy: with demographic adjustments, the proportion of variance predicted by neuropsychological measures declined significantly – by almost half – to a non-significant 11%. The expected pattern was seen for each of the five neuropsychological measures individually, although the difference was statistically significant only for CFT. This provides modest support for the generality of the detrimental effect of demographic corrections to the predictive relationship with driving across distinctive cognitive abilities.

Our findings are highly consistent with those of Silverberg and Millis (2009), providing an independent demonstration that when discrepancies between unadjusted and demographically-adjusted scores are present, unadjusted scores more accurately predict real-world functioning. Findings are also consistent with the observation of Reitan and Wolfson

(2005) that corrections for age are less useful for accurate interpretation of neuropsychological performances of a wide range of brain-damaged patients, for whom scores have diminished associations with age and education. Similarly, Golden and van den Broek (1998) reported that, among brain-damaged subjects, neuropsychological scores undergoing age-, education- and gender-corrections were less sensitive to impairment than were raw scores. They noted that by correcting for cognitive changes arising from aging, we may also be inadvertently adjusting for the effect of brain changes, "corrections" which make the individual appear normal when there may actually be subtle impairment.

Silverberg and Millis noted that the representativeness of their findings for populations with distinctive demographic characteristics and conditions other than TBI, and for functional capacities other than those included in their study, remained to be determined. The present study was designed to address the called for extensions by examining a distinctive outcome in a larger, demographically distinct (elderly, more highly educated and largely Caucasian), and clinically distinct sample with, or at risk for, mild cognitive decline. The AD and PD groups were selected for mildness of cognitive deficits for two reasons. First, individuals with mild cognitive decline are the most problematic in regards to determination of fitness for driving (Reger et al., 2004): There is ample evidence that patients with AD or PD *and marked cognitive impairment* are significantly worse drivers than healthy controls (Dawson et al., 2009a; Grace et al., 2005; Reger et al., 2004; Uc et al., 2009) and at greater risk for crashes (Lundberg et al., 1998; Stutts et al., 1998). However, even very mild cognitive decline may compromise safe driving (Withaar et al., 2000), and individuals with very mild dementia have been found to fail on-road driving tests at rates of 21% to 38% (Tallman et al., 1994, and Hunt et al., 1993, respectively). Thus, factors enhancing or detracting from accurate prediction of driving safety in individuals with very mild cognitive decline is of particular clinical import (Tallman, 1992). Second, inclusion of participants with more severe cognitive impairment would be expected to inflate findings regarding prediction with raw scores and diminished accuracy with demographically-adjusted scores, and it was decided that a stringent test would a more compelling demonstration of the hypothesized effect.

It is important to emphasize that, although the AD group was selected for mild cognitive decline, they did have significant cognitive impairments (most frequently in memory and associative verbal fluency), and all met standard, accepted diagnostic criteria for probable Alzheimer's. The MMSE cutoff was used strictly to identify those with very mild cognitive decline – not for diagnostic purposes as it has been established that the MMSE has limited sensitivity to very mild cognitive impairment in AD (Nasreddine et al., 2007), especially in a highly educated sample such as ours, and in PD as well (Hoops et al., 2009). In fact, patients with AD by rigorous diagnostic procedure can have MMSE scores of 30 (Shiroky et al., 2007).

Given the goal of this study, AD and PD groups were not intended to be representative of patients with AD and PD in general (i.e., with a broad range of disease severity), and we emphasize that this study was not designed to compare the driving ability of AD, PD and healthy elderly drivers. Rather, the follow-up analysis of the study groups was performed to examine the robustness of the superior predictive power of raw scores versus demographically-adjusted scores across neurological status. The results of that analysis showed that diminished predictive accuracy due to demographic adjustments was found for healthy elderly and one neurodegenerative disease (PD), but not for AD. Post hoc analyses revealed that, in the AD group, age was completely uncorrelated with driving errors ($r = .04$), in contrast to modest but non-negligible correlations among PD and NC ($r = .29$ and .27, respectively). It is possible that the restricted age range in this study and the generally later age of onset for AD contributed to this. In any event, the lack of association between

age and driving errors naturally minimizes the impact of correcting for age (the primary demographic correction in this study) on the multiple correlation between neuropsychological performances (adjusted scores) and driving errors for this group. Although this study was not designed to compare the driving ability of the various clinical conditions, it must be noted that AD and PD did not have significantly more driving errors than the controls, and this result warrants comment. In part, the lack of significance reflects the considerable variability in driving performance within each group. Additionally, compromised driving performance is not a function of diagnostic status per se but of the degree of cognitive weaknesses (Hunt et al., 1993; Tallman et al., 1994) and the inclusion criterion, designed to select a sample with very mild cognitive decline, accordingly yielded an AD group that does not reflect the severity of cognitive impairment found in AD with more advanced disease. It may be further noted that impairment in early AD generally involves memory most prominently, and memory tends to not be significantly correlated with driving performance (Dawson et al., 2009a).

Findings also extend the range of outcomes for which adjusted scores show diminished predictive accuracy, from ratings of broad functional capacities in the study of Silverberg and Millis (2009) (e.g., global functioning or employability) to a specific ability, driving, in the present study. In the earlier study, although the pattern of superiority for unadjusted scores was seen rather consistently across outcome measures, the magnitude of the discrepancy was weak for several measures. In the present study, the more impressive magnitude of the superiority for unadjusted scores (close to twice the predictive accuracy of demographically-corrected scores) may reflect the advantage conferred by the objective, reliable, ecologically valid outcome measure (DeRaedt & Ponjaert-Kristoffersen, 2001). This is consistent with the observation that neuropsychological tests are more strongly related to on-road driving tests than to caregiver reports of driving ability (Reger et al., 2005).

Although this study was not designed to evaluate or compare the predictive abilities of individual neuropsychological tests, it may be noted that the correlations between the neuropsychological measures (raw scores) – individually or in combination – and driving performance were in the low to moderate range (individually, $r$=.19–.38; collectively, .44). These correlations cannot be directly compared with findings from most previous studies due to (a) past studies' tendency to employ non-correlational analyses with dichotomous outcome measures (e.g., "safe" vs. "unsafe" driving), and (b) differing sample characteristics. Methodological differences notwithstanding, our results are consistent with previous findings. One review of on-road driving in older adults with cognitive decline (Withaar et al., 2000) found that such studies yield inconsistent results and show only low to moderate correlations between neuropsychological tests and on-road driving performance. In a more recent meta-analysis, Reger and colleagues (2005) converted results from 27 empirical studies into mean correlation coefficients weighted by sample size for each of six cognitive domains. Analyses of subjects with mild or very mild dementia showed that visuospatial skills bore the strongest relationship to on-road driving performance, a moderate weighted mean correlation of .29, followed by tests of attention/concentration, .25. Tests of mental status, language and executive functions showed small or essentially nonexistent effect sizes, .13, .10, and −.06 respectively. When control subjects were included, mean correlations were inflated, ranging from .48 (attention) to .36 (executive functions). Regarding specific tests, one recent study reported correlation coefficients (Whelihan et al., 2005), and it found that among patients with mild dementia TMT-A and TMT-B correlated with on-road driving performance .32 and .46, respectively. Additionally, their subjects were administered the Brief Visuospatial Memory Test, a very close relative of the BVRT employed in the present study, and they found learning to correlate .17 with driving. The generally low to moderate relationships (with the exception of occasionally

large effect sizes for the less widely used Useful Field of View Test of Ball and Owsley, 1993) may be somewhat disappointing. However, the present study was not designed to search for the optimal combination of neuropsychological measures, and the range of cognitive intactness/impairment was restricted, attenuating the correlations between cognitive scores and driving performances. Furthermore, from a broader perspective, the correlations from the present study are also consistent with the general level of univariate correlations that are typically found between cognitive measures and complex real-world functioning in areas such as job performance, job training performance, college GPA and graduate school GPA, which range from .19 to .28 (Meyer et al., 2001).

This study is not without limitations. Of the five neuropsychological measures, only age corrections were available for three (CFT, BVRT and BD). Although for the CFT normative data are available stratifying subjects by gender, education and age (Ardila et al., 1989), subjects in that study were Spanish speaking and age bands did not extend beyond age 65, making these norms unsuitable for the present study. The effect of the absence of gender and education corrections for these three measures is unknown; however, Silverberg and Millis (2009) found age to be the most powerful factor contributing to discrepancies with absolute scores, making it most relevant to the issue at hand. Additionally, they noted that, to the extent that multiple demographic adjustments increased a discrepancy from the absolute score, this tended to increase the reduction in predictive accuracy of demographically-adjusted scores. Thus, the unavailability of gender and education corrections for three predictors, if anything, likely attenuated the magnitude of the effect under study. That the demographic adjustments came from different sources for each test is another methodologic weakness as different normative samples may have differed from our study sample in divergent ways. However, the samples for the BVRT and CFT norms appear quite comparable to the study sample, and the TMT scores are corrected for each of the relevant factors, so samples differences should have minimal impact. The sample from which norms for the WAIS-III BD test were derived is less educated than the study sample. As this difference is not corrected for in the demographic adjustments for BD, it is likely that this also serves to underestimate the extent to which demographic adjustments would have reduced predictive accuracy had they included correction for education. In short, the limitation regarding the normative data available for some measures in the present study likely provided for a conservative test of the hypothesis.

Another possible limitation is the a priori restriction of the age range to 65 or older. It may be questioned whether a broader age range (e.g., age 60+) would have been more appropriate, especially since it was found that driving errors were not significantly related to age. This non-significant result is attributable, in part, to the sample's restricted age range. More importantly, that finding is consistent with previous findings that, *among elderly drivers*, impaired on-road driving performance is not an effect of aging per se, but of cognitive decline in a subset of that group (Fitten et al., 1995; Tallman et al., 1994). Neither of these observations threaten the internal validity of the study. The question is whether restricting the age range may have altered the primary study findings. The rationale for the restrictive age cutoff was to have a more stringent test of the study hypothesis. The powerful association of age, especially advanced age, to driving safety is well established (e.g., Massie et al., 1995; Williams & Shabanova, 2003). Given a sample with a broader age range and more powerful age effect in the raw data, the demographic adjustments – which are designed to eliminate the effect of age (and other demographic factors) – will necessarily inflate the extent to which demographic adjustments reduce predictive accuracy vis-à-vis raw scores.

Another limitation is the fact that driving ability was not assessed in terms of actual accidents and cited violations in day-to-day driving. It is possible that driving performance

was affected, for better or worse, by ways in which an on-road driving test differs from natural driving activities (e.g., a researcher sitting in the car evaluating them). However, history of accidents is also an imperfect criterion because they are very low-frequency events (Stutts et al., 1998); they are related to miles driven, and individuals with AD have been found to drive fewer miles (Massie et al., 1995; Trobe et al., 1996); and accidents are multidetermined so that some crashes are not necessarily attributable to an involved driver's ability (Lundberg et al, 1998; Withaar et al., 2000). On-road driving tests predict some types of at-fault crashes at levels comparable to the most predictive neuropsychological tests – including UFOV, the computerized neuropsychological test most highly correlated with crashes (Ball et al., 1993; Owsley et al., 1998) – and better than most (DeRaedt and Ponjaert-Kristoffersen, 2001), and at levels higher than caregiver-reported driving skill (Reger et al., 2005). Another concern is that use of an on-road driving test limits experimental control over potentially important variables, such as amount of traffic and the behavior of other drivers, and this may increase error variance in analyses. Nevertheless, active drivers do not drive in controlled settings, and on-road drives allow researchers to observe driving ability in real-world situations versus simulated drives in a controlled laboratory setting, increasing the ecological validity of study findings, an inherently valuable feature in research investigating important, complex real-world activities (Spooner and Pachana, 2006). In this study, driving errors were tallied according to well-specified criteria and were rated with good interrater reliability (.73), as is typical of such tests (DeRaedt and Ponjaert-Kristoffersen, 2001). Due to their advantages over actual crashes or caregiver-reported driving skill, on-road tests are generally considered the "gold standard" for assessing driving ability (Reger et al., 2005).

Demographic corrections are an important feature of neuropsychological assessment. However, present findings, in conjunction with those of Silverberg and Millis (2009), indicate that demographic corrections diminish predictive accuracy for complex real-world competencies. Findings also demonstrate that compromised driving safety is not a function of increasing age per se but of the degree to which relevant abilities are compromised in the individual driver. It is emphasized that this study was not designed to compare diagnostic groups, and so does not speak to the relative ability of drivers with AD, PD or healthy elderly. It will be important for future research to develop predictive equations without demographic corrections that incorporate measures of relevant neuropsychological abilities, and to empirically validate the levels of risk of impaired driving associated with various levels of performance. In conclusion, when the issue at hand is an individual examinee's ability to perform a complex real-world activity, it is their absolute level of ability that is pertinent, and basing recommendations on demographically-adjusted scores risks diminished accuracy of inferences regarding competency for that activity.

## Acknowledgments

## REFERENCES

American Academy of Clinical Neuropsychology Board of Directors. American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. The Clinical Neuropsychologist. 2007; 21:209–231. [PubMed: 17455014]

Ardila A, Rosselli M, Rosas P. Neuropsychological assessment in illiterates: Visuospatial and memory abilities. Brain and Cognition. 1989; 11:147–166. [PubMed: 2803759]

Ball K, Owsley C. The Useful Field of View Test: A new technique for evaluating age-related declines in visual function. Journal of the American Optometry Association. 1993; 64:71–79.

Ball, K.; Owsley, C. Predicting vehicle crashes in the elderly: Who is at risk?. In: Johansson, K.; Lundberg, C., editors. Aging and driving. Stockholm: Karolinska Institutet.; 1994. p. 115-128.

Benton AL, Eslinger PJ, Damasio AR. Normative observations on neuropsychological test performances in old age. Journal of Clinical Neuropsychology. 1981; 3:33–42. [PubMed: 7276195]

Dawson JD, Anderson SW, Uc EY, Dastrup E, Rizzo M. Predictors of driving safety in early Alzheimer's disease. Neurology. 2009a; 72:521–527. [PubMed: 19204261]

Dawson, JD.; Uc, EY.; Anderson, SW.; Dastrup, E.; Johnson, AM.; Rizzo, M. Ascertainment of On-Road Safety Errors Based on Video Review; Proceedings of Driving Assessment 2009: The Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design; Iowa City, Iowa: The University of Iowa; 2009b. p. 419-426.

DeRaedt R, Ponjaert-Kristoffersen I. Predicting at-fault car accidents of older drivers. Accident Analysis and Prevention. 2001; 33:809–818. [PubMed: 11579983]

Golden C, van den Broek A. Potential impact of age- and education-corrected scores on HRNB score patterns in participants with focal brain injury. Archives of Clinical Neuropsychology. 1998; 13:683–694. [PubMed: 14590628]

Grace J, Amick M, Abreu A, Festa E, Heindell W, Ott B. Neuropsychological deficits associated with driving performance in Parkinson's and Alzheimer's disease. Journal of the International Neuropsychological Society. 2005; 11:766–775. [PubMed: 16248912]

Fitten LJ, Perryman KM, Wilkinson CJ, Little RJ, Burns MM, Pachana N, Mervis JR, Malmgren R, Siembieda DW, Ganzell S. Alzheimer and vascular dementias and driving: A prospective road and laboratory study. Journal of the American Medical Association. 1995; 273:1360–1365. [PubMed: 7715061]

Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: A practical method of grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975; 12:189–198. [PubMed: 1202204]

Heaton, R.; Grant, I.; Matthews, C. Comprehensive norms for an expanded Halstead-Reitan neuropsychological battery: Demographic corrections, research findings, and clinical applications. Odessa, FL: Psychological Assessment Resources; 1991.

Heaton, RK.; Miller, SW.; Taylor, MJ.; Grant, I. Revised comprehensive norms for an expanded Halstead-Reitan battery: Demographically adjusted neuropsychological norms for African-American and Caucasian adults. Lutz, FL: Psychological Assessment Resources; 2004.

Hoops S, Nazem S, Siderowf AD, Duda JE, Xie SX, Stern MB, Weintraub D. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson's disease. Neurology. 2009; 73:1738–1745. [PubMed: 19933974]

Hunt L, Morris JC, Edwards D, Wilson BS. Driving performance in persons with mild senile dementia of the Alzheimer's type. Journal of the American Geriatrics Society. 1993; 41:747–753. [PubMed: 8315186]

Ivnik R. Normative psychology: A professional obligation? The Clinical Neuropsychologist. 2004; 19:159–161. [PubMed: 16019701]

Kleinbaum, DG.; Kupper, LL.; Nizam, A.; Muller, KE. Applied Regression Analysis and Multivariable Methods (4th ed.). London: International Thomason Publishing Inc.; 2008.

Lezak, MD.; Howieson, D.; Loring, D. Neuropsychological Assessment (4th ed.). New York: Oxford University Press; 2004.

Lundberg C, Hakamies-Blomqvist L, Almkvist O, Johansson K. Impairments of some cognitive functions are common in crash-involved older drivers. Accident Analysis and Prevention. 1998; 30:371–377. [PubMed: 9663296]

Massie DL, Campbell KL, Williams AF. Traffic accident involvement rates by driver age and gender. Accident Analysis and Prevention. 1995; 27:73–87. [PubMed: 7718080]

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's disease. Neurology. 1984; 34:939–944. [PubMed: 6610841]

Meyer G, Finn SE, Eyde LD, Kay GG, Moreland KL, Dies RR, Eisman EJ, Kubiszyn TW, Reed GM. Psychological testing and psychological assessment: A review of evidence and issues. American Psychologist. 2001; 56:128–165. [PubMed: 11279806]

Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society. 2007; 53:695–699. [PubMed: 15817019]

Owsley C, Ball K, McGwin G, Sloane ME, Roenker DL, White MF, Overley ET. Visual processing impairment and risk of motor vehicle crash among older adults. Journal of the American Medical Association. 1998; 279:1083–1088. [PubMed: 9546567]

Reger M, Welsh R, Watson G, Cholerton B, Baker L, Craft S. The relationship between neuropsychological functioning and driving ability in dementia: A meta-analysis. Neuropsychology. 2004; 18:85–93. [PubMed: 14744191]

Reitan R, Wolfson D. The effect of age and education transformations and neuropsychological test scores or persons with diffuse or bilateral brain damage. Applied Neuropsychology. 2005; 12:181–189. [PubMed: 16422659]

Rizzo, M.; McGehee, D.; Petersen, AD.; Dingus, TA. Development of an unobtrusively instrumented field vehicle for objective assessment of driving performance. In: Rothengatter, T.; Carbonnel, VE., editors. Traffic and Transport Psychology: Theory and Application. New York: Pergamon Press; 1997. p. 203-208.

Shiroky JS, Schipper HM, Bergman H, Chertkow H. Can you have dementia with an MMSE score of 30? American Journal of Alzheimer's Disease and Other DementiasCan you have dementia with an MMSE score of 30? American. Journal of Alzheimer's Disease and Other Dementias. 2007; 22:406–415.

Silverberg ND, Millis SR. Impairment versus deficiency in neuropsychological assessment: Implications for ecological validity. Journal of the International Neuropsychological Society. 2009; 15:94–102. [PubMed: 19128532]

Spooner DM, Pachana NA. Ecological validity in neuropsychological assessment: A case for greater consideration of research with neurologically intact populations. Archives of Clinical Neuropsychology. 2006; 21:327–337. [PubMed: 16769198]

Stutts JC, Stewart JR, Martel C. Cognitive test performance and crash risk an older driver population. Accident Analysis and Prevention. 1998; 30:337–346. [PubMed: 9663293]

Tallman, KS. Doctoral dissertation. Vancouver, British Columbia, Canada: University of British Columbia; 1992. Driving performance in mild dementia. (UMI No. NN80842)

Tallman, KS.; Beattie, BL.; Tuokko, H. The driving and aging study: Investigating the driving problems of those suffering from dementia. In: Johansson, K.; Lundberg, C., editors. Aging and driving. Stockholm: Karolinska Institutet.; 1994. p. 143-144.

Tranel, D. The Iowa-Benton school of neuropsychological assessment. In: Grant, I.; Adams, KM., editors. Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders. 3rd ed.. New York: Oxford University Press; 2009. p. 66-83.

Trobe JD, Waller PF, Cook-Flannagan CA, Teshima SM, Bieliauskas LA. Crashes and violations among drivers with Alzheimer disease. Archives of Neurology. 1996; 53:411–416. [PubMed: 8624215]

Uc EY, Rizzo M, Anderson SW, Sparks JD, Rodnitzky RL, Dawson JD. Impaired navigation in drivers with Parkinson's disease. Brain. 2007; 130:2433–2440. [PubMed: 17686809]

Uc EY, Rizzo M, Johnson AM, Dastrup E, Anderson SW, Dawson JD. Road Safety in Drivers with Parkinson Disease. Neurology. 2009; 73:2112–2119. [PubMed: 20018639]

Van Gorp WG, Satz P, Mitrushina M. Neuropsychological processes associated with normal aging. Developmental Neuropsychology. 1990; 6:279–290.

Wechsler, D. Wechsler Adult Intelligence Scale – Third edition: Administration and scoring manual. San Antonio, TX: Harcourt Brace & Co.; 1997.

Whelihan W, DiCarlo M, Paul R. The relationship of neuropsychological functioning to driving competence in older persons with early cognitive decline. Clinical Neuropsychology. 2005; 20:217–228.

Williams AF, Shabanova VI. Responsibility of drivers, by age and gender, for motor-vehicle crash deaths. Journal of Safety Research. 2003; 34:527–531. [PubMed: 14733986]

Withaar FK, Brouwer WB, Van Zomeren AH. Fitness to drive in older drivers with cognitive impairment. Journal of the International Neuropsychological Society. 2000; 6:480–490. [PubMed: 10902417]

**Table 1**

Demographic, Cognitive and Driving Characteristics of the Study Sample

| Variable | Full sample (N = 83) | Diagnostic Group | | | | Test Statistic[a] | Group Differences[e] |
|---|---|---|---|---|---|---|---|
| | | Controls (n = 24) | Alzheimer's (n = 26) | Parkinson's (n = 33) | | | |
| Gender | | | | | | 12.30** | |
| Male | 58 (69.9%) | 12 (50%) | 16 (61.5%) | 30 (90.9%) | | | |
| Female | 25 (30.1%) | 12 (50%) | 10 (38.5%) | 3 (0.9%) | | | |
| Age (yrs) | 72.1 (5.5) | 73.1 (5.1) | 75.2 (5.5) | 72.0 (4.1) | | 3.64* | |
| Education (yrs) | 15.7 (3.1) | 16.2 (2.7) | 16.2 (3.2) | 14.9 (3.2) | | 1.68 | |
| MMSE score | 28.6 (1.3) | 29.1 (1.3) | 28.5 (1.5) | 28.4 (1.2) | | 2.17 | |
| TMT-A | 45.9 (18.3) | 35.5 (8.6) | 53.3 (23.4) | 47.6 (15.6) | | 7.06** | c |
| TMT-B | 111.8 (54.0) | 77.3 (18.8) | 134.9 (69.9) | 118.8 (44.6) | | 9.07*** | c |
| CFT | 28.0 (4.1) | 29.1 (3.6) | 28.8 (3.9) | 26.5 (4.4) | | 3.51* | d |
| BVRT | 6.5 (3.4) | 3.8 (2.0) | 7.7 (3.2) | 7.5 (3.3) | | 14.00*** | c |
| BD | 32.0 (11.3) | 37.2 (11.2) | 30.7 (10.5) | 29.0 (10.9) | | 4.13* | c |
| Driving errors[e] | 39.8 (15.0) | 37.3 (13.8) | 40.0 (16.8) | 41.5 (14.4) | | 0.57 | |

*Note.* MMSE = Mini-Mental State Exam; TMT-A = Trail Making Test, Trail A; TMT-B = Trail Making Test, Trail B; CFT = Complex Figure Task Copy; BVRT = Benton Visual Retention Test Errors; BD = Wechsler Adult Intelligence Scale-III Block Design Test. Values reflect means and standard deviations except for gender. Follow-up tests were conducted with Fisher's protected (or LSD) test.

[a] Differences between study groups are tested by one-way analysis of variance (tests statistic, *F*), with the exception of gender which is evaluated by chi-square with two degrees of freedom.

[b] Between-group differences on neuropsychological scores were evaluated post hoc with Fisher's protected (LSD) test.

[c] AD, PD performed worse than NC.

[d] PD performed worse than NC, AD.

[e] Driving errors are as defined by the Iowa Department of Transportation for assessment of driving performance.

[†] *p* < .10.

[*] *p* < .05.

[**] *p* < .01.

***
$p < .001$.

**Table 2**

Bivariate Relationships between Neuropsychological Measures and Total Driving Errors

| Test | Raw scores Pearson $r$ | Adjusted scores Pearson $r$ | $Z^a$ |
|---|---|---|---|
| TMT-A | .258[*] | .219[*] | 0.75 |
| TMT-B | .187[†] | .152 | 0.58 |
| CFT | .384[***] | .207[†] | 3.56[***] |
| BVRT | .206[†] | .177 | 0.48 |
| BD | .269[*] | .207[†] | 1.64 |

*Note*. TMT-A = Trail Making Test, Trail A; TMT-B = Trail Making Test, Trail B; CFT = Complex Figure Task Copy; BVRT = Benton Visual Retention Test Errors; BD = Wechsler Adult Intelligence Scale-III Block Design Test. For all tests for which lower scores indicate higher functioning, the direction of the correlations was reversed so that a positive correlation always indicates a positive relationship between better neuropsychological performance and better driving performance.

[a] $Z$ tests the statistical significance of the difference between the correlation coefficients associated with raw scores vs. adjusted scores.

[†] $p < .10$.

[*] $p < .05$.

[**] $p < .01$.

[***] $p < .001$.