Comparing Caregiver and Clinician Predictions of Fitness to Drive in People With Alzheimer’s Disease

Kimberly Bixby, Jennifer D. Davis, Brian R. Ott

This observational study investigated family caregiver and clinician ratings of 75 drivers with Alzheimer’s disease against scores on a standardized road test and a naturalistic driving evaluation. Clinician ratings by a physician specialized in dementia were significantly associated with road test error scores ($r = .25$, $p = .03$) but not naturalistic driving errors or global ratings of road test and naturalistic driving performance. Caregiver ratings were unrelated to either driving assessment, with two exceptions; adult child ratings of driving ability were correlated with road test error scores ($r = .43$, $p = .02$), and spousal ratings were inversely correlated with global ratings. Clinician ratings of driving competence were modestly correlated with road test performance, but caregiver ratings were more complex. Adult children may be more accurate reporters of driving ability than spouses, possibly because of less personal bias, but the reasons behind this discrepancy need further investigation.


长大了于65岁的人占美国总人口的13%和所有许可驾驶者的16%（National Center for Statistics and Analysis, 2013）。老年人面临许多可能导致危险驾驶的问题，包括视力问题、与年龄相关的身体衰退以及认知能力问题。这项研究集中于认知能力下降作为危险驾驶的主要风险因素。估计表明，三分之一的痴呆症患者虽然有认知障碍，仍继续开车（Silverstein et al., 2011）。之前的研究已经证明，痴呆症患者作为整体，患有轻度阿尔茨海默病的老年人是更高风险的司机，而且随着痴呆症恶化，驾驶能力变差（Dubinsky, Stein, & Lyons, 2000; Ott, Papandonatos, Davis, & Barco, 2012）。驾驶是一个复杂的活动，要求整合和协调多种认知技能，影响程度由退行性疾病决定，如记忆、视觉和信息处理、执行功能、反应时间和注意力。因为老年痴呆症的进展性，所有患有这种疾病的司机最终必须停止驾驶。

尽管许多患有AD的司机在诊断后仍然开车，但没有标准化的措施可以评估驾驶特定技能的驾驶者，其认知障碍。标准化路测被视为监测老年驾驶者驾驶能力的标准工具（Carr & Ott, 2010）。标准化路测表现与可获得的办公室评估和历史记录中的车辆碰撞关系良好（Barrash et al., 2010; Brown, Stern, et al., 2005; Dawson, Anderson, Uc, Dastrup, & Rizzo, 2009; Fitten et al., 1995; Grace et al., 2005）。标准化路测也是估计老年人驾驶能力的一种合理近似，特别是在日常生活驾驶环境中（Davis et al., 2012; Ott et al., 2012）。

虽然标准化路测有局限性，可能会很昂贵，并且许多家庭有限制享受这种评估。因此，自我报告和
caregiver report measures that clinicians can use in the office to identify driving errors and risky behaviors are desirable. Several studies have found limited utility for self-report measures in AD because affected people experience reduced insight into their own cognitive deficits and denial of their reduced driving performance (Brown, Ott, et al., 2005; Cotrell & Wild, 1999; Dobbs, 1997; Hunt, Morris, Edwards, & Wilson, 1993; O’Neill et al., 1992; Wild & Cotrell, 2003). Families and caregivers often rely on input from physicians to make the final decision about driving cessation. In one study comparing participant, caregiver, and neurologist ratings of driving ability on a 3-point Likert scale (safe, marginal, unsafe) to ratings by a professional driving instructor, only the neurologist’s ratings related to the standardized road test score (Brown, Ott, et al., 2005). Further studies demonstrated that agreement between physician and driving instructor ratings of driving fitness ranged between 62% and 78% (Ott et al., 2005, 2012). In another study, a physician’s predictions of participants’ standardized road test performance correlated with road test score \((r = .63)\) but not global impression of driving performance (Fox, Bowden, Bashford, & Smith, 1997). Taken together, these studies indicate that valid and reliable questionnaires regarding driving errors are needed in the office setting for early screening and possible detection of reduced driving skills.

Relatively little research has been done to understand caregivers’ accuracy in predicting driving ability in people with AD. Studies using global ratings of driving fitness found that although caregivers were more stringent than people with AD, their ratings did not relate significantly to performance on a standardized road test given by a professional driving instructor (Brown, Ott, et al., 2005; Hunt et al., 1993). Studies using specific questionnaires to assess caregiver reports of driving skills and behaviors, such as the self-report Safe Driving Behavior Measure, have shown significant agreement between driving evaluators and caregivers (Classen et al., 2012a, 2012b). Although caregiver ratings may be more reliable predictors of driving performance than self-report ratings, they generally are not as accurate as a clinician’s global impression. For this reason, the American Academy of Neurology recommended that caregiver ratings of driving be considered lower level support for driving cessation in its practice guidelines (Iverson et al., 2010).

The goal of our study was to evaluate the accuracy of clinician and caregiver ratings of driving fitness compared with a standardized measure of road test performance and extended on-road video recordings of naturalistic driving behavior. Previous reports have shown fair agreement between global ratings on road test and naturalistic driving evaluation and a significant association between error scores in the two assessments and similar error types (Davis et al., 2012). For this study, we defined a caregiver as a spouse or adult child who was in routine contact and drove at least monthly with the memory-impaired participant. We hypothesized that because caregivers are more familiar with their family member’s driving in routine environments, caregiver ratings would correlate more closely with naturalistic driving performance than with performance on a standardized road test. In addition, we explored caregiver relationship to the driver (i.e., spouse vs. adult child) as a potential modifying variable in the caregiver ratings.

Method

Study Design

We conducted an observational study of performance on a formal, standardized on-road test and naturalistic driving behavior in older adults with cognitive impairment.

Participants

Participants were 75 drivers with mild to moderate AD who were ages 60–89 yr, were English speaking, currently drove at least one trip per week, and had a valid driver’s license (validated by visual inspection of the license at the office visit). Participants were being treated for AD in a hospital-based memory disorders center and were recruited during routine follow-up visits. All participants met diagnostic criteria for possible or probable AD based on criteria established by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (McKann et al., 1984). Participants had Clinical Dementia Rating Scale (CDR; Morris, 1993) ratings of very mild \((0.5; n = 35)\), mild \((1.0; n = 33)\), or moderate \((2.0; n = 5)\) dementia.

For each participant, a caregiver reported information on driving habits (e.g., number of trips, miles driven per week) and driving skills. Only drivers with either a spouse \((n = 36)\) or adult child \((n = 27)\) caregiver were included in this study to examine potential differences between spouse and adult child reports. Family members were required to have spent time with the participant at least once weekly and to have accompanied the study participant while driving at least one monthly during the preceding 12 mo.

Exclusion criteria included ophthalmological, physical, or neurological disorders other than dementia that would impair driving ability; reversible dementia; intellectual disability; schizophrenia; bipolar disorder; and alcohol or substance abuse within the previous year. Participants were...
instructed to abstain from use of alcohol, minor and major tranquilizers, and narcotics for at least 24 hr before study visits and road tests. In addition, participants’ dosages of symptomatic antidementia drugs (e.g., cholinesterase inhibitors, gingko) and substances that could influence disease progression (e.g., Vitamin E, estrogen, nonsteroidal anti-inflammatory drugs) must have been stable for at least 6 wk before study entry.

All participants provided written informed consent. The Rhode Island Hospital institutional review board approved all procedures.

**Procedures**

**Office Visit.** Participants had a brief in-office clinical assessment that included a neurological examination, a vision screen, the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), and the CDR (Morris, 1993). All participants’ neurological examination results were considered consistent with AD. A research assistant conducted a semistructured interview with each family member and participant to collect driving information such as history of driving collisions and violations over the previous 3 yr, miles driven per week, trips per week, and miles driven with the caregiver per week.

**Clinician and Caregiver Ratings.** During the office visit, each family member and a neurologist rated the participants’ global driving abilities on a trichotomous scale of 0 = drives alone without difficulty (safe), 1 = drives alone, but with some difficulty (marginal), or 2 = drives with some difficulty and only if accompanied (possibly unsafe). The neurologist based his ratings on information obtained during the initial diagnostic evaluation. Family caregivers also made yes—no ratings of 12 driving skills on a more detailed questionnaire of specific driving behaviors such as running stop signs, speeding, and near misses (Brashear, Unversagt, & Kuhn, 2002); responses provide a frequency rating of types of driving error, with higher scores reflecting more error types. Caregivers and the neurologist were blinded to both standardized road test and naturalistic driving outcomes.

**Standardized Road Test.** Within 1 mo after the office visit, a professional driving instructor trained in the evaluation of neurologically disabled drivers administered the Rhode Island Road Test (RIRT) to participants. The driving instructor was blinded to the driver’s diagnosis and office evaluation results. The RIRT was modeled after the Washington University Road Test and was adapted for comparable streets in Rhode Island (Hunt et al., 1997). Details of the administration and scoring of the RIRT have been previously published (Davis et al., 2012; Hunt et al., 1997). The RIRT has established reliability and validity (Brown, Ott, et al., 2005; Davis et al., 2012).

The RIRT covers 6.5 miles of urban terrain without highway driving and requires 45 min to complete during daylight hours and under good road conditions. The instructor provides drivers only with oral instructions to complete the course. The instructor rates 28 driving maneuvers on a 3-point Likert scale (0 = unimpaired, 1 = mildly impaired, and 2 = moderately to severely impaired). Example maneuvers include making right and left turns; changing lanes; negotiating signage and traffic lights; and correctly using mirrors, steering wheel, and other vehicle equipment. Total scores range from 0 to 960, and an error severity score is calculated by dividing the sum of ratings for each event by the number of observed maneuvers; as such, the RIRT error score is a proportion, with higher scores reflecting poorer performance.

The driving instructor also makes a categorical rating of overall driving ability: pass, pass with recommendations, marginal with restrictions, marginal with training, or fail. A participant passes the road test if the driving instructor judges the risk for errors leading to a crash to be low; this rating indicates safe driving behavior. A marginal rating indicates that the driver can continue to drive but should restrict driving to certain times and locations or should enroll in driving or retraining lessons. A fail rating indicates that the driving behaviors create a high risk for future crashes. High-risk behaviors include instrument and pedal confusion or consistent failure to respond to signage or traffic lights. For this study, global ratings were collapsed into three categories: (1) pass (combined pass and pass with recommendations), (2) marginal (combined marginal with restrictions and marginal with training), and (3) fail.

**Video-Recorded Naturalistic Driving.** Of the 75 participants who took the RIRT, 59 passed and had cameras installed in their vehicles to record 2 wk of naturalistic driving along their normal routes. The details of administration and scoring of video recordings of driving have previously been described (Davis et al., 2012; Ott et al., 2012). Four low-profile cameras (3.5 in. × 2.5 in. × 3.0 in.) recorded driving from the front and back dashboards to give views of the driver and all around the vehicle. One front camera faced the participant, another faced the road, and the back two cameras faced outward to capture the environment to the sides of the vehicle. Participants followed their daily routine and drove in their normal environment. A research assistant reviewed all video recordings, and the driving instructor reviewed approximately 4 hr of daytime driving. Nighttime driving was not coded because of suboptimal visibility.

The driving instructor rated naturalistic driving using the Composite Driving Assessment Scale (CDAS), a scale
developed by authors Ott and Davis. Interrater agreement for the CDAS was established in a prior study (Ott et al., 2012), and reliabilities fell in the moderate range for the global ratings ($κ = .45$) and error scores ($r = .62$). The instructor assigned a global rating to specific driving behaviors, including discrete events (maneuvers) and global events (attention, attitude, reaction time) using the same 3-point Likert scale as the RIRT (0 = unimpaired, 1 = mildly impaired, and 2 = moderately to severely impaired). The ratings reflected the driver’s overall skill level for each behavior across a sample of 4 hr of recorded driving. The total score ranged from 0 to 60. Given that the driving instructor was viewing maneuvers across 4 hr, an average severity rating was calculated for each maneuver (i.e., sum of ratings for each event divided by number of observed maneuvers). The instructor also provided a categorical rating (pass, marginal, or fail) of overall driving ability.

**Statistical Analyses**

We compared baseline characteristics between groups using independent-sample $t$ tests for continuous variables and $χ²$ for categorical variables. We tested the relationship between road test and naturalistic driving error scores and global ratings and caregiver and clinician ratings using biserial and Spearman’s $ρ$ correlations. Correlations were interpreted using guidelines recommended by Portney and Watkins (2009).

**Results**

Participants’ average age was 76.60 yr ($SD = 6.32$); 41% had a high school education or less, and 57% had some college education. Dementia severity was generally mild, with average MMSE scores of 25.05 ($SD = 2.84$). Participants had been driving on average 56.28 ($SD = 9.23$) yr, were taking 9.11 ($SD = 6.22$) trips per week, and were driving 87.62 ($SD = 88.22$) miles per week. The majority of the drivers were male (53.3%) and White (92.0%).

The average error severity rating was .09 ($SD = .07$) for the RIRT and .19 ($SD = .13$) for naturalistic driving. Table 1 presents the driving instructor’s global ratings on the RIRT and CDAS and the clinician and caregiver ratings of driving ability.

**Clinician Ratings of Driving Ability**

Clinician ratings were significantly but poorly associated with road test error scores ($r = .25$, $p = .03$) but not with global ratings ($ρ = .07$, $ρ = .55$). Clinician ratings were not significantly correlated with naturalistic driving error scores ($r = .16$, $p = .24$) or global ratings ($ρ = .15$, $p = .26$).

**Caregiver Ratings of Driving Ability**

Table 2 shows the caregivers’ relationship to the drivers. On average, caregivers spent 79.15 ($SD = 70.22$) hr/wk with the driver. With regard to driving exposure, caregivers drove an average of 20.33 ($SD = 41.50$) miles and took 1.98 ($SD = 3.26$) trips per week with the driver. Compared with adult children, spouses spent many more hours with the driver per week and spent much more time in the car directly observing driving (see Table 2).

Overall, global caregiver ratings were not significantly related to the driving instructor’s road test error scores ($r = .22$, $p = .08$) or global ratings ($ρ = −.10$, $ρ = .46$); caregiver ratings also were not significantly related to naturalistic driving error scores ($r = .16$, $p = .25$) or global ratings ($ρ = .17$, $ρ = .23$). When relationship was considered, adult child global ratings were moderately related to road test error scores ($r = .43$, $p = .02$) but not to naturalistic driving error scores ($r = .08$, $p = .73$) or to global ratings on the road test ($ρ = .31$, $ρ = .12$) or naturalistic driving evaluation ($ρ = .22$, $ρ = .33$). Spousal ratings were not significantly related to road test error scores ($r = .00$, $p = .99$), but they were inversely related to global ratings on the road test ($ρ = −.47$, $ρ = .01$), with better spousal ratings being associated with poorer global ratings by the driving instructor. We found no significant correlations between spousal ratings and naturalistic driving outcomes ($r = .02$, $p = .99$; $ρ = .13$, $ρ = .51$).

Caregivers also provided dichotomous ratings of specific driving maneuvers in which they endorsed a driving

**Table 1. Ratings of Driving Safety by Driving Instructor, Clinician, and Caregivers**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Global Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road test rating by driving instructor ($n = 76$)</td>
<td>Pass Marginal Fail</td>
</tr>
<tr>
<td>Naturalistic driving rating by driving instructor ($n = 59$)</td>
<td>21 12 26</td>
</tr>
<tr>
<td>Rater</td>
<td>Safe Marginal Possibly Unsafe</td>
</tr>
<tr>
<td>Clinician ($n = 73$)</td>
<td>17 38 18</td>
</tr>
<tr>
<td>Caregiver ($n = 64$)</td>
<td>36 23 5</td>
</tr>
</tbody>
</table>

**Table 2. Caregiver Exposure to Participants’ Driving, by Relationship**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Spouse ($n = 36$)</th>
<th>Adult Child ($n = 27$)</th>
<th>$t$ or $x^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hr with driver/wk</td>
<td>133.68 (43.54)</td>
<td>16.14 (31.97)</td>
<td>11.73 .000</td>
<td></td>
</tr>
<tr>
<td>Trips/wk</td>
<td>3.58 (3.99)</td>
<td>0.41 (0.70)</td>
<td>4.08 .000</td>
<td></td>
</tr>
<tr>
<td>Miles/wk</td>
<td>36.63 (52.89)</td>
<td>3.17 (5.62)</td>
<td>3.21 .000</td>
<td></td>
</tr>
<tr>
<td>Sex (% female)</td>
<td>25</td>
<td>78</td>
<td>18.96 .000</td>
<td></td>
</tr>
</tbody>
</table>

*Note. M = mean; SD = standard deviation.*
skill as a problem or not (yes—no response). Correlations between the specific scaled items and error scores in both the naturalistic driving and road test assessments are presented in Table 3. For the adult children, the sum of ratings was positively correlated with both road test error scores ($r = .48$, $p = .02$) and naturalistic driving error scores ($r = .54$, $p = .03$). When considering specific driving maneuvers, adult child ratings for difficulty finding vehicle controls ($r = .49$, $p = .02$; $r = .65$, $p < .001$), driving too slowly on streets ($r = .50$, $p = .02$; $r = .68$, $p = .01$), and driving too slowly on interstates ($r = .71$, $p < .001$; $r = .69$, $p = .02$) were positively correlated with road test and naturalistic driving error scores, respectively. In contrast, the sum of spouses’ overall ratings was unrelated to road test errors ($r = -.36$, $p = .13$) and negatively correlated with naturalistic driving error scores ($r = -.56$, $p = .04$).

**Discussion**

Consistent with prior work, we found that clinician ratings were significantly but weakly associated with road test performance (Brown, Ott, et al., 2005) but that their association with naturalistic driving performance did not achieve significance. In contrast, caregiver ratings were poor indicators of both naturalistic driving and road test performance; these results were contrary to our hypothesis that caregiver ratings would be associated with naturalistic driving scores because of caregivers’ real-world contact with the drivers. Further investigation into caregiver ratings, however, showed that adult children’s global ratings and ratings of specific driving behaviors were more closely related to road test performance than those of spouses. Surprisingly, spouses rated their family members’ driving abilities in the wrong direction; better spousal ratings were associated with poorer objective naturalistic driving scores.

The reason underlying the poor association between caregiver and objective ratings is unclear and was not empirically tested in this study. A number of reasons have been put forth to explain reporter bias in caregivers, including the desire to avoid interpersonal conflict or to deemphasize the progression of the disease because of reliance on the driver for transportation (Wild & Correll, 2003). In a study of 80 older drivers, 33% stated that their independence would be affected if their spouse stopped driving, even though they were all licensed drivers (Classen et al., 2012a). In addition, because caregiving for a person with Alzheimer’s disease requires providing support in many activities of daily living, spouses may not see poor driving skills as being as problematic as other dementia-related behaviors (Pinquart & Sørensen, 2011). Last, spouses may not emphasize problems in transportation because they may view the decline in functioning as a failure in their caregiving abilities (Taylor & Tripodes, 2001).

Similarly, adult children may underreport their parent’s ability to drive because having a parent dependent on them for transportation might disrupt their daily functioning. A survey of 315 households with older drivers found that many caregivers reported that they had missed work (42%) or had stopped working completely (13%) when the care recipient became unable to drive; the vast majority of these caregivers were adult children (Taylor & Tripodes, 2001).

**Table 3. Caregiver Report of Participants’ Driving Skills, by Relationship**

<table>
<thead>
<tr>
<th>Item</th>
<th>Correlation With Error Scores</th>
<th>Road Test</th>
<th>Naturalistic Driving</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spouse Adult Child</td>
<td>Spouse Adult Child</td>
<td></td>
</tr>
<tr>
<td>Gets lost while driving</td>
<td>−.49*</td>
<td>.16</td>
<td>−.34</td>
</tr>
<tr>
<td>Runs red lights</td>
<td>−.35</td>
<td>.20</td>
<td>−.40</td>
</tr>
<tr>
<td>Runs stop signs</td>
<td>−.33</td>
<td>−.14</td>
<td>−.49</td>
</tr>
<tr>
<td>Has trouble staying in the lane or crosses over the center line</td>
<td>.13</td>
<td>.38</td>
<td>−.30</td>
</tr>
<tr>
<td>Has trouble finding the controls in the vehicle</td>
<td>−.28</td>
<td>.49*</td>
<td>−.50</td>
</tr>
<tr>
<td>Goes the wrong way on a one-way street</td>
<td>−.23</td>
<td>−.18</td>
<td>−.23</td>
</tr>
<tr>
<td>Goes the wrong way on the interstate*</td>
<td></td>
<td></td>
<td>−.01</td>
</tr>
<tr>
<td>Drives too slowly on a city street</td>
<td>−.35</td>
<td>.50*</td>
<td>−.40</td>
</tr>
<tr>
<td>Drives too slowly on the interstate</td>
<td>−.35</td>
<td>.71**</td>
<td>−.32</td>
</tr>
<tr>
<td>Speeds on a city street</td>
<td>−.20</td>
<td>.03</td>
<td>−.47</td>
</tr>
<tr>
<td>Speeds on the interstate</td>
<td>−.13</td>
<td>−.05</td>
<td>−.17</td>
</tr>
<tr>
<td>Hits or almost hits another vehicle, person, or object</td>
<td>−.35</td>
<td>−.07</td>
<td>−.67**</td>
</tr>
<tr>
<td>Total</td>
<td>−.36</td>
<td>.48*</td>
<td>−.56</td>
</tr>
</tbody>
</table>

*No caregiver endorsed “goes the wrong way on the interstate,” so no correlations were calculated for this item.

*p < .05, **p < .01.
Our data suggest that adult children are better reporters of driving skills than spouses, but we did not collect data on the possible negative consequences for the adult child’s life if his or her parent stopped driving.

This study has several important limitations. Recruitment was completed by convenience sampling, which limited the ethnic diversity of our sample and potential generalizability of our findings to families of non-White ethnic backgrounds. In addition, all participants were drawn from a memory clinic with reliable caregivers who were willing to participate in both road tests and video observation of naturalistic driving; thus, our sample may have comprised a more compliant group of clients with less risky driving habits. Although our sample had a range of driving abilities, only drivers deemed fit to drive by a driving instructor were allowed to have cameras installed in their cars. Consequently, we were unable to evaluate associations between the naturalistic driving ratings of clinicians and caregivers of the most unfit drivers. We were also unable to examine how the clinician weighed family report of driving concerns in his driving rating because he was blind to all assessments in the study. Future studies may examine whether responses on family-reported driving questionnaires influence clinicians’ accuracy in predicting driving fitness.

Overall, caregiver ratings of driving ability should be used with caution when assessing people with AD, particularly when the caregiver is a spouse. Our results suggest that the value of caregiver predictions of driving skills could be improved by using a detailed questionnaire that addresses specific driving errors rather than global ratings; such surveys include the questionnaire developed by the American Academy of Neurology (Iverson et al., 2010) or the self-report Safe Driving Behavior Measure developed by Classen and colleagues (2012a, 2012b).

Implications for Occupational Therapy Practice

The results of this study have the following implications for occupational therapy practice:

- Use of a valid and reliable questionnaire of driving ability that addresses specific, rather than global, ratings of driving behaviors may provide valuable information when making decisions about fitness to drive with older adult clients with AD.
- The validity of family member reports of driving ability may be weak, but in general adult children may be better informants than spouses. Gathering input from multiple family members may be beneficial when completing a driving evaluation.
- Caregiver education programs could include training programs to help caregivers, especially spouses, recognize and monitor unsafe driving behaviors.

Acknowledgments

We thank Lindsay Miller, Scott Hewitt, and Stephanie Knott for their efforts in data collection; Timothy Souza of the ABC/ACE Driving School for completing the driving evaluations; and Peggy Barco for her review of prior versions of this article. This study was supported by the National Institute on Aging Grant No. R01 AG016335 to Brian R. Ott. The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References


